

59847

STIC-Biotech/Ch mLib

From: Chan, Christina  
Sent: Thursday, February 07, 2002 11:19 AM  
T : Chen, Shin-Lin  
Cc: STIC-Biotech/ChemLib  
Subject: RE: Rush sequence search: 09/391,606

CRFE

Please rush. Thanks Chris

-----Original Message-----

Fr m: Chen, Shin-Lin  
Sent: Wednesday, February 06, 2002 6:03 PM  
To: Chan, Christina  
Cc: STIC-Biotech/ChemLib  
Subject: Rush sequence search: 09/391,606

09/391,606 is an amended case due next bi-week and require sequence search. Could you approve the following Rush Sequence Search for 09/391,606? Thanks!

SEQ ID Nos. 1-4, 7-9 and 12-16.

*Shin-Lin Chen*

AU 1632  
CM1 12A15  
Mail Box # 12E12  
(703)305-1678

POINT OF CONTACT:  
BARB O'BRYEN  
TECH. INFORMATION SPECIALIST  
STIC CM1 ~~12014~~ 308-4291

6A05

Searcher: Barb  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 2-8-02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:34:38 ; Search time 137.02 Seconds  
(without alignments)  
212.997 Million cell updates/sec

Title: US-09-391-606-15  
Perfect score: 2103  
Sequence: 1 MLPVGNPDPSPDLLIDTWE.....QKLISEEDLN SAVDHHHHH 394

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID2/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	1925	91.5	391	20	AAV53319	Chlamydia pneumoniae
2	1647	78.3	343	20	AAV56771	C. trachomatis ser
3	1534.5	73.0	389	20	AAW98188	Chlamydia psittaci
4	1408	67.0	402	20	AAW98189	Chlamydia psittaci
5	1408	67.0	525	21	AAW13645	C. pneumoniae sero
6	1408	67.0	525	22	AAW32213	Protein encoded by
7	1407	66.9	402	20	AAW98187	Chlamydia psittaci
8	1356.5	64.5	387	20	AAV56767	C. trachomatis ser
9	1349	64.1	404	20	AAV56768	C. trachomatis ser
10	1345.5	64.0	376	19	AAW76366	C. trachomatis JMI
11	1338.5	63.6	397	20	AAV56766	C. trachomatis ser

12	1336.5	63.6	397	20	AAV56763	C. trachomatis ser
13	1326	63.1	396	20	AAV56764	C. trachomatis ser
14	1321.5	62.8	397	20	AAV56765	C. trachomatis ser
15	1320.5	62.8	376	19	AAW76363	C. trachomatis JMI
16	1313.5	62.5	393	20	AAV56757	C. trachomatis ser
17	1312.5	62.4	393	20	AAV56759	C. trachomatis ser
18	1306.5	62.2	372	19	AAW76365	C. trachomatis JMI
19	1306.5	62.2	393	22	AAE06646	Chlamydia trachoma
20	1306.5	62.1	374	19	AAW76364	C. trachomatis JMI
21	1305.5	62.1	395	20	AAV56762	C. trachomatis ser
22	1300	61.8	373	19	AAW76362	C. trachomatis JMI
23	1295	61.6	394	18	AAW15149	Chlamydia trachoma
24	1295	61.6	394	19	AAW73141	Chlamydia trachoma
25	1295	61.6	394	21	AAW81268	Chlamydia trachoma
26	1294	61.5	394	19	AAW57775	Chlamydia trachoma
27	1293	61.5	394	20	AAV56761	C. trachomatis ser
28	1289	61.3	392	20	AAV56760	C. trachomatis ser
29	1287	61.2	394	20	AAV56758	C. trachomatis ser
30	1284	61.1	394	20	AAW37494	Chlamydia trachoma
31	1280	60.9	394	7	AAW60004	Sequence of a major
32	1264.5	60.1	343	20	AAV56769	C. trachomatis ser
33	1248	59.3	457	21	AAW82388	C. trachomatis MOM
34	1214.5	57.8	484	21	AAW82389	C. trachomatis MOM
35	1182	56.2	356	20	AAV56770	C. trachomatis ser
36	1134.5	53.9	514	21	AAW82391	C. trachomatis MOM
37	1085.5	51.6	454	21	AAW82390	C. trachomatis MOM
38	1053	50.1	349	21	AAW82392	C. trachomatis MOM
39	913	43.4	228	20	AAW98184	Chlamydia psittaci
40	907	43.1	225	20	AAW98186	Chlamydia psittaci
41	842.5	40.1	222	20	AAW98183	Chlamydia psittaci
42	825.5	39.3	215	20	AAW98185	Chlamydia psittaci
43	767	36.5	277	21	AAW82393	C. trachomatis MOM
44	519	24.7	100	20	AAW95307	Chlamydia major o
45	489	23.3	100	20	AAW95295	Chlamydia major o

## ALIGNMENTS

RESULT 1  
RAY35319  
ID AAY35319 standard; Protein; 391 AA.  
XX AAY35319;  
XX  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE Chlamydia pneumoniae transmembrane protein sequence.  
XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
XX  
OS Chlamydia pneumoniae.  
XX  
PN WO9927105-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98WO-IB01890.  
XX  
PR 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.  
XX (GIST ) GENSET.  
XX  
PI Griffiths R;  
XX  
DR WPI; 1999-357842/30.  
XX  
PT Genome sequence of Chlamydia pneumoniae  
XX  
PS Page 1130-1131; Disclosure; 1912pp; English.



XX AAY34584-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX  
SQ Sequence 391 AA;

Query Match 91.5%; Score 1925; DB 20; Length 391;  
Best Local Similarity 98.9%; Pred. No. 2.3e-182;  
Matches 362; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 61  
DB LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 85  
QY 62 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVFCILGASNG 121  
DB SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVFCILGASNG 145  
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYDTFSFSWVGARGALWECGC 181  
DB YIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYDTFSFSWVGARGALWECGC 205  
QY 182 ATLGAEFQYAOQKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKSAI 241  
DB ATLGAEFQYAOQKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKSAI 265  
QY 242 NYHEWQVQASLSYRLNSLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAWNPSSLGN 301  
DB NYHEWQVQASLSYRLNSLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAWNPSSLGN 325  
QY 302 ATALSTTDSFDFMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 361  
DB ATALSTTDSFDFMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 385  
QY 362 SQGFRF 367  
DB SQGFRF 391

RESULT 2  
AAY56771  
ID AAY56771 standard; Protein; 343 AA.

XX AAY56771;

XX 22-FEB-2000 (first entry)

XX C. trachomatis serovar HuPu MOMP sequence.

XX Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;  
XX cellular response; immunogen; Th1-like CD4 response; mucosal immunity.

XX Chlamydia trachomatis.

XX W09951745-A2.

XX 14-OCT-1999.

XX 07-APR-1999; 99WO-CA00292.

XX 07-APR-1998; 98US-0055765.

XX (UYMA-) UNIV MANITOBA.

XX Bruhnam RC;  
PI WPI; 1999-620205/53.  
DR  
XX  
PT Non-replicating vector encoding fragments of the outer membrane protein  
PT of Chlamydia, useful in vaccines and as immunogen  
XX  
PS Disclosure; Fig 10 A-F; 52pp; English.  
XX  
CC The invention provides a non-replicating vector that comprises, linked  
CC to a promoter, a nucleotide sequence that encodes a region containing at  
CC least one of the conserved domains 2, 3 and 5 of a major outer membrane  
CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in  
CC vaccines to generate a protective immune response (mainly cellular)  
CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful  
CC in standard immunoassays. Immunization with the vector induces a broad  
CC spectrum of immune responses, including Th1-like CD4 responses and  
CC mucosal immunity, providing significant protection against subsequent  
CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety  
CC of serovars of C. trachomatis.  
XX  
SQ Sequence 343 AA;

Query Match 78.3%; Score 1647; DB 20; Length 343;  
Best Local Similarity 86.3%; Pred. No. 7.2e-155;  
Matches 316; Conservative 2; Mismatches 2; Indels 46; Gaps 1;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 61  
DB LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 83  
QY 62 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVFCILGASNG 121  
DB SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVFCILGASNG 143  
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYDTFSFSWVGARGALWECGC 181  
DB YIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYDTFSFSWVGARGALWECGC 193  
QY 182 ATLGAEFQYAOQKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKSAI 241  
DB ATLGAEFQYAOQKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKSAI 217  
QY 242 NYHEWQVQASLSYRLNSLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAWNPSSLGN 301  
DB NYHEWQVQASLSYRLNSLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAWNPSSLGN 277  
QY 302 ATALSTTDSFDFMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 361  
DB ATALSTTDSFDFMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 337  
QY 362 SQGFRF 367  
DB SQGFRF 343

RESULT 3

AAW98188  
ID AAW98188 standard; Protein; 389 AA.

XX AAW98188;

XX 05-JUL-1999 (first entry)

XX Chlamydia psittaci major outer membrane protein.

XX Major outer membrane protein; MOMP; psittacosis; infection;  
XX vaccine; genetic immunisation.

XX Chlamydia psittaci.

XX



PN WO9910005-A1.  
 XX 04-MAR-1999.  
 XX 28-AUG-1998; 98WO-US17943.  
 XX 28-AUG-1997; 97US-0057147.  
 XX (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.  
 PA Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
 PI WPI; 1999-254214/21.  
 XX DR N-PSDB; AAX25047.  
 XX A new vaccine for Chlamydia psittaci infections  
 XX Disclosure; Page 60-61; 72pp; English.  
 XX The present sequence is the major outer membrane protein (MOMP)  
 CC of Chlamydia psittaci strain B577. A claimed MOMP polypeptide (see  
 CC AA98184) comprises regions V03 and V04 of B577 MOMP, i.e. it lacks  
 CC regions V01 and V02. A claimed vaccine composition includes MOMP  
 CC polypeptide lacking V01 and V02, optionally fused to a maltose  
 CC binding protein. Also claimed are an isolated nucleic acid  
 CC encoding the polypeptide, a vector, and a method of preventing C.  
 CC psittaci infection by administering the vaccine containing the  
 CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking  
 CC regions V01 and V02 are useful for genetic vaccination. The  
 CC vaccines are used to prevent C. psittaci infection, especially in  
 CC birds.  
 XX Sequence 389 AA;  
 SQ  
 Query Match 73.0%; Score 1534.5; DB 20; Length 389;  
 Best Local Similarity 75.7%; Pred. No. 1.3e-143;  
 Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;  
 QY 2 LPVGNPSPSLIDGTIWEGAGDPCDPCATWCDATSLRAGFYGYVDFRILKVDAPKTF 61  
 Db 23 LPVGNPAEPLIDGTWEGASGDCPCSTWCDATSLRAGFYGYVDFRILKVDVNTKI 82  
 QY 62 S-MGAKP--TGSAA-ANYTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVFCCTL 120  
 Db 83 TGMGAVPTGTAANYKTPTDRPNIAVGHKLQDAEFTNAAFLALNIWDRFDFICTLGASN 142  
 QY 121 GYIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYTDTSFSNSVGARGALWECG 180  
 Db 143 GYFKASSAFLVGLIGVKGSSIAADQIPNVGITGIVGYVDFRILKVDVNTKI 202  
 QY 181 CATLGAEFQYAGSKPKVEELNVCNVSQFVNKPKGYGVAPPLPTDAGVATATGTSAT 240  
 Db 203 CATLGAEFQYAGSKPKVEELNVCNVSQFVNKPKGYGVAPPLPTDAGVATATGTSAT 262  
 QY 241 INYHEWQVGCASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLTAVNLTAWNPISLIG 300  
 Db 263 IKYHEWQVGLAFLYRLNLVPLVYISVNWSRATFDADNIRIAQPKLTAVNLTAWNPISLIG 322  
 QY 301 NATALSTDSFDFMQIVSCQINKFKSRKACGVTVGTATLVADKWSLTAEARLINERAH 360  
 Db 323 EATALDTSNKFDALFIQASIQINKMKSRKACGVTVGTATLVADKWSLTAEARLINERAH 382  
 QY 361 VSGQFRF 367  
 Db 383 MNAQFRF 389  
 RESULT 4  
 AA98189  
 ID AA98189 standard; Protein; 402 AA.  
 XX  
 AC AA98189;

XX 05-JUL-1999 (first entry)  
 DT Chlamydia psittaci major outer membrane protein.  
 XX Major outer membrane protein; MOMP; psittacosis; infection;  
 DE vaccine; genetic immunisation.  
 KW Chlamydia psittaci.  
 XX OS WO9910005-A1.  
 XX PN 04-MAR-1999.  
 XX PD 28-AUG-1998; 98WO-US17943.  
 XX PF 28-AUG-1997; 97US-0057147.  
 XX PR (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.  
 XX PA Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
 XX PI WPI; 1999-254214/21.  
 XX DR N-PSDB; AAX25048.  
 XX A new vaccine for Chlamydia psittaci infections  
 PT Disclosure; Page 65-66; 72pp; English.  
 PS The present sequence is the major outer membrane protein (MOMP)  
 XX of Chlamydia psittaci strain 6BC. Claimed MOMP polypeptides (see  
 CC AA98183 and AA98184) comprise regions V03 and V04 of an MOMP, i.e.  
 CC they lack regions V01 and V02. Claimed vaccine compositions  
 CC include such MOMP polypeptides, optionally fused to a maltose  
 CC binding protein. Also claimed are isolated nucleic acids encoding  
 CC the polypeptide, a vector, and a method of preventing C. psittaci  
 CC infection by administering the vaccine containing the MOMP  
 CC polypeptide. Vectors encoding MOMP polypeptides lacking regions  
 CC V01 and V02 are useful for genetic vaccination. The vaccines are  
 CC used to prevent C. psittaci infection, especially in birds.  
 XX Sequence 402 AA;  
 SQ  
 Query Match 67.0%; Score 1408; DB 20; Length 402;  
 Best Local Similarity 70.3%; Pred. No. 4.7e-131;  
 Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;  
 QY 2 LPVGNPSPSLIDGTIWEGAGDPCDPCATWCDATSLRAGFYGYVDFRILKVDAPKTF 61  
 Db 23 LPVGNPAEPLIDGTWEGASGDCPCSTWCDATSLRAGFYGYVDFRILKVDVNTKI 82  
 QY 62 S-MGAKP--TGSAA-ANYTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVFCCTL 116  
 Db 83 TGMGAVPTGTAANYKTPTDRPNIAVGHKLQDAEFTNAAFLALNIWDRFDFICTLGASN 142  
 QY 117 GASNGYIRGNSTAFNLVGLFGVKG---TTVNANLPSVLSNGVVELYTDTSFSNSVGA 172  
 Db 143 GASNGYIRGNSTAFNLVGLFGVKG---TTVNANLPSVLSNGVVELYTDTSFSNSVGA 202  
 QY 173 RGALEWCCATLGAEFQYAGSKPKVEELNVCNVSQFVNKPKGYGVAPPLPTDAGVATATGTSAT 230  
 Db 203 RGALEWCCATLGAEFQYAGSKPKVEELNVCNVSQFVNKPKGYGVAPPLPTDAGVATATGTSAT 262  
 QY 231 ATATGTSATINYHEWQVGCASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLTAVNL 290  
 Db 263 EATALDTSNKFDALFIQASIQINKMKSRKACGVTVGTATLVADKWSLTAEARLINERAH 322  
 QY 291 LTAWNPISLIG 347  
 ID AA98189 standard; Protein; 402 AA.  
 XX  
 AC AA98189;







Query Match 67.0%; Score 1408; DB 22; Length 525;  
 Best Local Similarity 70.3%; Pred. No. 7, 1e-131;  
 Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPDSLLIDGTWEGAGDPCDPCATWCDALISLRAGFYGVDFRILKVDAPKTF 61  
 Db 146 lpgvnpaepslldgtwewegagdpdpcatwcdalsiragvygdyvdrvlkvdvntkf 205

QY 62 S-MGAKPT---GSAA--ANTTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFDFVCTL 116  
 Db 206 sgmaatptcaginasntnqpeangrpnlaygrhmqdaefnsaafialniwdrfdlfccl 265

QY 117 GASNGYIRGNSTAFNLVGLFVGK-----TTVNANELPNVSLNGVVELYDTSTFSWSVGA 172  
 Db 266 gasngyfkassaafnlvlgfsaassistdipmqlpnvgitqgvvfevfytdtsfswsvga 325

QY 173 RGALWECGCATGAEFYAQSKPKVEELNVICNVSFVSNKPKYKGA--PPLPTDAGV 230  
 Db 326 rgawecgcatagaefyagsqnpkiemlnvtspaqfvlhkpgrkgassnfpplptagt 385

QY 231 ATATGKTSATINHEWQVGSLSYRLNLSLPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290  
 Db 386 teatdtktsatikyheqwglalsyrlnmlypvgvnwratfdadtiriaqpkkseiln 445

QY 291 LTAWNPSLLGNATAL---STTDSFSDFMQIVSCQINKFKSRKACGVTVGATLVADKWSL 347  
 Db 446 ittwnpsllgstaltalpnngskdvlgsdlqiasiqinkmksrkacgvavgatliadadkwsl 505

QY 348 TAEARLINERAHVSGQRF 367  
 Db 506 tgearlneraahmaaqfrf 525

RESULT 7  
 AAW98187  
 ID AAW98187 standard; Protein; 402 AA.  
 XX AAW98187;  
 AC AAW98187;  
 DT 05-JUL-1999 (first entry)  
 DE Chlamydia psittaci major outer membrane protein.  
 KW Major outer membrane protein; MOMP; psittacosis; infection;  
 KW vaccine; genetic immunisation.  
 XX Chlamydia psittaci.  
 OS WO9910005-A1.  
 PN 04-MAR-1999.  
 PD 28-AUG-1998; 98WO-US17943.  
 PF 28-AUG-1997; 97US-0057147.  
 PR (LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.  
 XX Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
 PI WPI; 1999-254214/21.  
 DR N-PSDB; AAX25046.  
 XX A new vaccine for Chlamydia psittaci infections  
 PT Disclosure; Page 56-57; 72pp; English.  
 PS The present sequence is the major outer membrane protein (MOMP)  
 CC of Chlamydia psittaci strain LSUWCK, a cockatiel isolate (the  
 CC MOMP gene sequence of this isolate is identical to that of C.  
 CC psittaci Avian Type C). A claimed MOMP polypeptide (see AAW98183)  
 CC comprises regions VD3 and VD4 of LSUWCK MOMP, i.e. it lacks  
 CC regions VD1 and VD2. A claimed vaccine composition includes MOMP

CC polypeptide lacking VD1 and VD2, optionally fused to a maltose  
 CC binding protein. Also claimed are an isolated nucleic acid  
 CC encoding the polypeptide, a vector, and a method of preventing C.  
 CC psittaci infection by administering the vaccine containing the  
 CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking  
 CC regions VD1 and VD2 are useful for genetic vaccination. The  
 CC vaccines are used to prevent C. psittaci infection, especially in  
 CC birds.

XX Sequence 402 AA;  
 SQ

Query Match 66.9%; Score 1407; DB 20; Length 402;  
 Best Local Similarity 70.0%; Pred. No. 6e-131;  
 Matches 266; Conservative 39; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPDSLLIDGTWEGAGDPCDPCATWCDALISLRAGFYGVDFRILKVDAPKTF 61  
 Db 23 lpgvnpaepslldgtwewegagdpdpcatwcdalsiragvygdyvdrvlkvdvntkf 82

QY 62 S-MGAKPT---TGSAA--ANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFDFVCTL 116  
 Db 83 sgmaatptcaginasntnqpeangrpnlaygrhmqdaefnsaafialniwdrfdlfccl 142

QY 117 GASNGYIRGNSTAFNLVGLFVGK-----TTVNANELPNVSLNGVVELYDTSTFSWSVGA 172  
 Db 143 gasngyfkassaafnlvlgfsaassistdipmqlpnvgitqgvvfevfytdtsfswsvga 202

QY 173 RGALWECGCATGAEFYAQSKPKVEELNVICNVSFVSNKPKYKGA--PPLPTDAGV 230  
 Db 203 rgawecgcatagaefyagsqnpkiemlnvtspaqfvlhkpgrkgassnfpplptagt 282

QY 231 ATATGKTSATINHEWQVGSLSYRLNLSLPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290  
 Db 263 teatdtktsatikyheqwglalsyrlnmlypvgvnwratfdadtiriaqpkkseiln 322

QY 291 LTAWNPSLLGNATAL---STTDSFSDFMQIVSCQINKFKSRKACGVTVGATLVADKWSL 347  
 Db 323 ittwnpsllgstaltalpnngskdvlgsdlqiasiqinkmksrkacgvavgatliadadkwsl 382

QY 348 TAEARLINERAHVSGQRF 367  
 Db 383 tgearlneraahmaaqfrf 402

RESULT 8  
 AAV56767  
 ID AAV56767 standard; Protein; 387 AA.  
 XX AAV56767;  
 AC AAV56767;  
 DT 22-FEB-2000 (first entry)  
 DE C. trachomatis serovar MoPn MOMP sequence.  
 KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;  
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.  
 XX Chlamydia trachomatis.  
 OS WO9951745-A2.  
 PN 14-OCT-1999.  
 PD 07-APR-1999; 99WO-CA00292.  
 PF 07-APR-1998; 98US-0055765.  
 PR (UYMA-) UNIV MANITOBA.  
 XX Bruhnam RC;  
 XX WPI; 1999-620205/53.







KW Major outer membrane protein; antibody; antigen; diagnosis; disease.  
 XX Chlamydia trachomatis.  
 OS JPI0234395-A.  
 XX  
 PN 08-SEP-1998.  
 XX  
 PD 25-FEB-1997; 97JP-0040780.  
 XX  
 PF 25-FEB-1997; 97JP-0040780.  
 XX  
 PR (ELED ) DENKI KAGAKU KOGO KK.  
 XX  
 PA WPI; 1998-535045/46.  
 XX  
 DR Preparation of major outer membrane protein of Chlamydia trachomatis  
 XX - by cloning and recombinant expression of the gene, for use as a  
 PT diagnostic of Chlamydia infections  
 XX  
 PS Claim 2; Page 15-16; 19pp; Japanese.  
 XX  
 CC AAM76362-W76366 are major outer membrane proteins isolated from  
 CC Chlamydia trachomatis which are used for the measurement of an antibody  
 CC against Chlamydia trachomatis, by using it as an antigen in the form of  
 CC a reagent. The method can give a diagnostic method for Chlamydia  
 CC trachomatis-infected diseases with high specificity.  
 XX  
 SQ Sequence 376 AA;

Query Match 64.0%; Score 1345.5; DB 19; Length 376;  
 Best Local Similarity 65.2%; Pred. No. 6.8e-125;  
 Matches 245; Conservative 51; Mismatches 71; Indels 9; Gaps 3;

Qy 1 MLPVGNPSLLIDGTIWEAGAGDPCDCAISLRAGFYGDYVDFRLKVDAPKT 60  
 Db 1 MLPVGNPAEPSIMIDGILWEGFGGDPDPCATWCDALSMRVYGGDFRVLKDVNKE 60

Qy 61 FSMGAKPTGSAANY-----TTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVCTL 116  
 Db 61 fmggaaptndaaadlqndpknvarpnpaygkqhmqdaemftnaaymalniwdrfvcctl 120

Qy 117 GASNGYIRGNSTAFNLVGLFGVKGTTVNAE---LPNVSLNGVVELYDTDFSFWSVGAR 173  
 Db 121 gattgylkgnasnlvlgftkssdfntaklvpnlalnrvavelydttdtfawsvgar 180

Qy 174 GALWECGCATLCAEFOYQAKSPKVEELNVICNVQSFVNKPKGYKGVAPPLPTDAGVATA 233  
 Db 181 aalwecgcatalgasfygaqskpkveelnvlnaseftinkpkgygaefpdlitagtea 240

Qy 234 TGTKSATINYHEWQVGSLSRLNSLVPIYGVQWSRATFDADNIRIAQPKLPTAVLNLT 293  
 Db 241 tgtkdsidyhewqaslsalrylnmftpygvkwsrvsfadtiriaqpklaealldvtt 300

Qy 294 WNPSSLGNAT--ALSTTDSFDMQIVSCQINKFKSRKACGVTVGCATLVADKWSLTAE 351  
 Db 301 lnptiagktvvasgsdndladtmqivslqlnkmksrksclavgttividadkyavtvet 360

Qy 352 RLINERAAHVSGQFRF 367  
 Db 361 rlideraahvnaqfrf 376

RESULT 11  
 AAY56766  
 ID AAY56766 standard; Protein; 397 AA.  
 XX  
 AC AAY56766;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE C. trachomatis serovar H MOMP sequence.

XX Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;  
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.  
 XX Chlamydia trachomatis.  
 OS WO9951745-A2.  
 XX  
 PN 14-OCT-1999.  
 XX  
 PD 07-APR-1999; 99WO-CA00292.  
 XX  
 PF 07-APR-1998; 98US-0055765.  
 XX  
 PR (UYMA-) UNIV MANITOBA.  
 XX  
 PA Bruhnam RC;  
 XX  
 DR WPI; 1999-620205/53.  
 XX  
 PT Non-replicating vector encoding fragments of the outer membrane protein  
 XX of Chlamydia, useful in vaccines and as immunogen  
 XX  
 PS Disclosure; Fig 10 A-F; 52pp; English.  
 XX  
 CC The invention provides a non-replicating vector that comprises, linked  
 CC to a promoter, a nucleotide sequence that encodes a region containing at  
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane  
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in  
 CC vaccines to generate a protective immune response (mainly cellular)  
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful  
 CC in standard immunoassays. Immunization with the vector induces a broad  
 CC spectrum of immune responses, including Th1-like CD4 responses and  
 CC mucosal immunity, providing significant protection against subsequent  
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety  
 CC of serovars of C. trachomatis.  
 XX  
 SQ Sequence 397 AA;

Query Match 63.6%; Score 1338.5; DB 20; Length 397;  
 Best Local Similarity 64.8%; Pred. No. 3.7e-124;  
 Matches 243; Conservative 52; Mismatches 71; Indels 9; Gaps 3;

Qy 2 LPVGNPSLLIDGTIWEAGAGDPCDCAISLRAGFYGDYVDFRLKVDAPKTF 61  
 Db 23 lpgvnpaepsimidgilwegfggdpdpcatwcdalsmrvggygdfvrlktdvnkef 82

Qy 62 SMGAKPTGSAANY-----TTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVCTLG 117  
 Db 83 qmggaaptndaaadlqndpknvarpnpaygkqhmqdaemftnaaymalniwdrfvcctlg 142

Qy 118 ASNGYIRGNSTAFNLVGLFGVKGTTVNAE---LPNVSLNGVVELYDTDFSFWSVGARG 174  
 Db 143 attgylkgnasnlvlgftkssdfntaklvpnlalnrvavelydttdtfawsvgara 202

Qy 175 ALWECGCATLCAEFOYQAKSPKVEELNVICNVQSFVNKPKGYKGVAPPLPTDAGVATAT 234  
 Db 203 alwecgcatalgasfygaqskpkveelnvlnaseftinkpkgygaefpdlitagtea 262

Qy 235 GTKSATINYHEWQVGSLSRLNSLVPIYGVQWSRATFDADNIRIAQPKLPTAVLNLTAW 294  
 Db 263 gtkdsidyhewqaslsalrylnmftpygvkwsrvsfadtiriaqpklaealldvttl 322

Qy 295 NPSLLGNAT--ALSTTDSFDMQIVSCQINKFKSRKACGVTVGCATLVADKWSLTAEAR 352  
 Db 323 nptiagktvvasgsdndladtmqivslqlnkmksrksclavgttividadkyavtvetr 382

Qy 353 LINERAAHVSGQFRF 367  
 Db 383 llderaahvnaqfrf 397



RESULT 12  
 AAY56763  
 ID AAY56763 standard; Protein; 397 AA.  
 XX  
 AC AAY56763;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE C. trachomatis serovar L3 MOMP sequence.  
 XX  
 KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;  
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN WO9951745-A2.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 07-APR-1999; 99WO-CA00292.  
 XX  
 PR 07-APR-1998; 98US-0055765.  
 XX  
 PA (UYMA-) UNIV MANITOBA.  
 XX  
 PI Bruhnam RC;  
 XX  
 DR WPI; 1999-620205/53.  
 XX  
 PT Non-replicating vector encoding fragments of the outer membrane protein  
 PT of Chlamydia, useful in vaccines and as immunogen  
 XX  
 PS Disclosure; Fig 10 A-F; 52pp; English.  
 XX  
 CC The invention provides a non-replicating vector that comprises, linked  
 CC to a promoter, a nucleotide sequence that encodes a region containing at  
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane  
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in  
 CC vaccines to generate a protective immune response (mainly cellular)  
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful  
 CC in standard immunoassays. Immunization with the vector induces a broad  
 CC spectrum of immune responses, including Th1-like CD4 responses and  
 CC mucosal immunity, providing significant protection against subsequent  
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety  
 CC of serovars of C. trachomatis.  
 XX  
 SQ Sequence 397 AA;

Query Match 63.6%; Score 1336.5; DB 20; Length 397;  
 Best Local Similarity 64.8%; Pred. No. 5.8e-124;  
 Matches 243; Conservative 50; Mismatches 73; Indels 9; Gaps 3;  
 QY 2 LPVGNPSPDLLIDGTWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 Db 23 lpvgnpaeplsimidgilwegf9gdpctctwcdaismrvygydfvdrvlktdvnef 82  
 QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFVFCFLG 117  
 Db 83 qmgaepstsdtaglsnptntvarpnaygkhdmdaemftnaaymalniwdrfvdvctlg 142  
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGKTVNANE---LPNVSLNSGVVELYTDTSFSSVSGARG 174  
 Db 143 attgylkgnasfnlvgifgtktstnfntaklvpntalnqavveltydttdttfawsvgara 202  
 QY 175 ALWEGCATLCAEFOYQASQKPKVEELNVCNVSQFSVKNPKGYGVAFPLPTDAGVATAT 234  
 Db 203 alwecgcatlgasfgyaqsqkpkveelnvldcasetfinkpkgygvaefpltditagteat 262  
 QY 235 GTSKATINHEWQVGSALSYRLNSLVPYIGVQWSRATFDADNIRIAQKPLPTAVNLNTAW 294  
 Db 263 gtdkasidyhewqsalysrlnmftpyigvkwsvsfadtdiriaqpklaeavldvttl 322

QY 295 NPSLL--GNATALSTTDSFDMQIVSQINKEKSRKAGCTVGTGATLVADKWSLTARAR 352  
 Db 323 nptiagkgsvvsgseneladtmqivslqlnkmksrkgciavgtcttvdadkyavivetr 382  
 QY 353 LINERAAHVSGQFRF 367  
 Db 383 llderaahvnaqfrf 397  
 RESULT 13  
 AAY56764  
 ID AAY56764 standard; Protein; 396 AA.  
 XX  
 AC AAY56764;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE C. trachomatis serovar A MOMP sequence.  
 XX  
 KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;  
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN WO9951745-A2.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 07-APR-1999; 99WO-CA00292.  
 XX  
 PR 07-APR-1998; 98US-0055765.  
 XX  
 PA (UYMA-) UNIV MANITOBA.  
 XX  
 PI Bruhnam RC;  
 XX  
 DR WPI; 1999-620205/53.  
 XX  
 PT Non-replicating vector encoding fragments of the outer membrane protein  
 PT of Chlamydia, useful in vaccines and as immunogen  
 XX  
 PS Disclosure; Fig 10 A-F; 52pp; English.  
 XX  
 CC The invention provides a non-replicating vector that comprises, linked  
 CC to a promoter, a nucleotide sequence that encodes a region containing at  
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane  
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in  
 CC vaccines to generate a protective immune response (mainly cellular)  
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful  
 CC in standard immunoassays. Immunization with the vector induces a broad  
 CC spectrum of immune responses, including Th1-like CD4 responses and  
 CC mucosal immunity, providing significant protection against subsequent  
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety  
 CC of serovars of C. trachomatis.  
 XX  
 SQ Sequence 396 AA;

Query Match 63.1%; Score 1326; DB 20; Length 396;  
 Best Local Similarity 64.7%; Pred. No. 6.3e-123;  
 Matches 242; Conservative 48; Mismatches 76; Indels 8; Gaps 3;

QY 2 LPVGNPSPDLLIDGTWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 Db 23 lpvgnpaeplsimidgilwegf9gdpctctwcdaismrvygydfvdrvlktdvnef 82  
 QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFVFCFLG 117  
 Db 83 qmgaepstsdtaglsnptntvarpnaygkhdmdaemftnaaymalniwdrfvdvctlg 142  
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGKTVNANE---LPNVSLNSGVVELYTDTSFSSVSGARG 174  
 Db 143 attgylkgnasfnlvgifgtktstnfntaklvpntalnqavveltydttdttfawsvgara 202



QY 1 MLPVGNPSPDLLIDGTIWEGAAGDPCDPCATWCDAISLRAGFYGDYVFDRILKVDPK 60



```

Db      1 mlpvgnpaepsImidgiiwegfggdpcttwcdaismrvggygdfvdrvlktdvke 60
Qy      61 FSMGAKPTGSAANY----TTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFDVFCFL 116
Db      61 fmggaapttsdvaglpnqoptinvarpnpaykqhmqdaemftnaaymalniwdrfadvctl 120
Qy      117 GASNGYIRGNSTAFNLVGLFGVKGTTVANE---LPNVSLNSGVVVELYTDTSFSWSVGAR 173
Db      121 gattgylkgnaasfnlvgfktqssfnaklipntalneavvelyinttfawsvgar 180
Qy      174 GALWECGCATLGAETQYAGSKPKVEELNVICNVQFSVKNPKGYKGVAFPLPTDAGVATA 233
Db      181 aalwecgcaticgasfgyagskpkveelnvlnaseftinkpgyvgaeftpinitagteaa 240
Qy      234 TGTKSATINYHEWQYGASLSYRLNSLVPIYGWQSRATFDADNIRIAOPKLPATAVLNLTA 293
Db      241 tgtkdasidnewqaslsalsyrlmmftpygkwsrvsfadtfiragpklaealldvt 300
Qy      294 WNPSSLGNATALS--TTDSFDFMOIVSCQINKFKSRKACGVTVGATLVADADKWSLTAEA 351
Db      301 lnrttagkgsvvsagtdneladtmqivslqlnkmksrkscglavgttividadkyavtvea 360
Qy      352 RLINERAAHVSCQPRF 367
Db      361 rliideraahvnaqfrf 376

```

Search completed: February 7, 2002, 21:34:40  
Job time: 20759 sec







FILING DATE: 14-AUG-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/915,503  
 FILING DATE: 14-AUG-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/US97/17618  
 FILING DATE: 01-OCT-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/US97/17885  
 FILING DATE: 01-OCT-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Apple, Randolph Ted  
 REGISTRATION NUMBER: 36,429  
 REFERENCE/DOCKET NUMBER: 015389-002610US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 611:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1154 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FEATURE:  
 NAME/KEY: Protein  
 LOCATION: 1..1154  
 OTHER INFORMATION: /note= "fusion protein composed of hprt  
 protein sequence, vector sequences, the  
 OTHER INFORMATION: Myc epitope and His6 tag"  
 US-08-974-549A-611

Query Match 6.8%; Score 142.5; DB 4; Length 1154;  
 Best Local Similarity 26.1%; Pred. No. 9.3e-06;  
 Matches 54; Conservative 19; Mismatches 37; Indels 97; Gaps 7;  
 QY 200 LNVICNVSO--FSVKNPK-----GYKGVAFPLPTDAGVATATGKTSATINHYHEQVVGAS 251  
 DB 1033 LAVISDTASLCYSILKAKNAGSLGAKGAAGPLPSEA-----VQWLCHQA 1077  
 QY 252 LSYRLN-----SLVPYGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLST 307  
 DB 1078 FLKLTRHRTVYVPLLG-----SLRTAQTL----- 1103  
 QY 308 TDSFDEMQIVSCQINKFKSRKACGVTVGATLVLDADKWSLTAERLINERAAHVSGQFRF 367  
 DB 1104 -----SRKLPTTLTAL-----EAAANPALPSDF 1127  
 QY 368 RYQAYVEQKLISEEDLNSAVDHHHHH 394  
 DB 1128 KTLIDLEQKLISEEDLNSAVDHHHHH 1154  
 RESULT 2  
 US-08-374-560-1  
 ; Sequence 1, Application US/08374560  
 ; Patent No. 5882645  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TOTH, Istvan  
 ; APPLICANT: GIBBONS, William Anthony  
 ; TITLE OF INVENTION: PEPTIDE COMPOUNDS  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
 ; STREET: 1180 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: US  
 ; ZIP: 10036-8403  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/374,560  
 FILING DATE: 13-MAR-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9215780.9  
 FILING DATE: 24-JUL-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mellman, Edward A.  
 REGISTRATION NUMBER: 24,735  
 REFERENCE/DOCKET NUMBER: P/365-302  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 382-0700  
 TELEFAX: (212) 382-0888  
 TELEX: 236925  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 61 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal  
 US-08-374-560-1  
 Query Match 6.5%; Score 136.5; DB 2; Length 61;  
 Best Local Similarity 47.5%; Pred. No. 3.4e-07;  
 Matches 29; Conservative 10; Mismatches 21; Indels 1; Gaps 1;  
 QY 269 RATEPDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLSTTD-SFSDFMQIVSCQINKFKS 327  
 DB 1 RASFDADTIRIAQPKSAETIFDVTTLNPTIAGAGDVKTSAGSQLGDTMQIVSLQLNKMK 60  
 QY 328 R 328  
 DB 61 R 61  
 RESULT 3  
 PCT-US93-08739-3  
 ; Sequence 3, Application PC/TUS9308739  
 ; GENERAL INFORMATION:  
 ; APPLICANT: The Government of the United States of America  
 ; APPLICANT: as represented  
 ; APPLICANT: by the Secretary of the Department of Health and Human Services  
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR  
 ; TITLE OF INVENTION: CHLAMYDIA  
 ; TITLE OF INVENTION: TRACHOMATIS  
 ; NUMBER OF SEQUENCES: 3  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/08739  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/947,671 US  
 ; FILING DATE: 18 SEP 93  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 42 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO



FRAGMENT TYPE: N-terminal  
PCT-US93-08739-3

Query Match 5.98; Score 125; DB 5; Length 42;  
Best Local Similarity 61.9%; Pred. No. 3.1e-06;  
Matches 26; Conservative 4; Mismatches 2; Indels 10; Gaps 2;

Qy 103 ALNIWDRFDVFTLGASNGYIRGNSAFNLVGLGVKGTTVN 144  
|||||  
Db 1 ALNIWDRFDVFTLGATGTYLKGN-----FDV--TTLN 32

## RESULT 4

US-08-446-692-45  
Sequence 45, Application US/08446692

Patent No. 5759551

## GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Y1

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 45:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-446-692-45

Query Match 5.98; Score 124; DB 1; Length 25;  
Best Local Similarity 84.0%; Pred. No. 1.7e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRFDVFTLGASNGYIRGNS 127  
|||||

Db 1 ALNIWDRFDVFTLGATGTYLKGN 25

## RESULT 5

US-08-488-351A-45

Sequence 45, Application US/08488351A

Patent No. 5843446

## GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Y1

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,351A

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/229,275

FILING DATE: 14-APR-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/057,166

FILING DATE: 27-APR-1992

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 45:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-488-351A-45

Query Match 5.98; Score 124; DB 2; Length 25;  
Best Local Similarity 84.0%; Pred. No. 1.7e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRFDVFTLGASNGYIRGNS 127  
|||||

Db 1 ALNIWDRFDVFTLGATGTYLKGN 25

## RESULT 6

PCT-US93-08739-1

Sequence 1, Application PC/TUS9308739

## GENERAL INFORMATION:

APPLICANT: The Government of the United States of America

APPLICANT: as represented

APPLICANT: by the Secretary of the Department of Health and Human Services

TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR

TITLE OF INVENTION: CHLAMYDIA

TITLE OF INVENTION: TRACHOMATIS

NUMBER OF SEQUENCES: 3

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:



APPLICATION NUMBER: PCT/US93/08739  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION NUMBER: 07/947,671 US  
FILING DATE: 18 SEP 93  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
PCT-US93-08739-1

Query Match 5.9%; Score 124; DB 5; Length 25;  
Best Local Similarity 84.0%; Pred. No. 1.7e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127  
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

## RESULT 7

PCT-US95-13841-17  
Sequence 17, Application PC/TUS9513841  
GENERAL INFORMATION:

APPLICANT: United Biomedical Inc; Walfield, Alan M.;  
APPLICANT: Wang, Chang Yi

TITLE OF INVENTION: Synthetic IgE Membrane Anchor  
TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy

NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10154

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/13841  
FILING DATE: 25-OCT-1995

CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/328,519  
FILING DATE: 25-OCT-1994

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: Lin, Maria C.H.  
REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4117  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849

TELEX: 421792  
INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-13841-17

Query Match 5.9%; Score 124; DB 5; Length 25;  
Best Local Similarity 84.0%; Pred. No. 1.7e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127  
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

## RESULT 8

US-09-411-329C-22  
Sequence 22, Application US/09411329C

Patent No. 6261820  
GENERAL INFORMATION:

APPLICANT: Boone, Thomas  
APPLICANT: Li, Huimin

APPLICANT: Mann, Michael  
TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE

FILE REFERENCE: A-596  
CURRENT APPLICATION NUMBER: US/09/411,329C

CURRENT FILING DATE: 1999-10-01  
NUMBER OF SEQ ID NOS: 29

SOFTWARE: PatentIn version 3.0  
SEQ ID NO 22

LENGTH: 27  
TYPE: PRT

ORGANISM: Agkistrodon contortrix  
NAME/KEY: misc\_feature

FEATURE:  
OTHER INFORMATION: Native pro-fibrolase of Agkistrodon contortrix

US-09-411-329C-22

Query Match 5.9%; Score 124; DB 4; Length 27;  
Best Local Similarity 87.5%; Pred. No. 1.9e-06;  
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 371 AYVEQKLISEDLNSAVDHHHHH 394  
Db 4 SFLEQKLISEDLNSAVDHHHHH 27

## RESULT 9

US-09-411-329C-26  
Sequence 26, Application US/09411329C

Patent No. 6261820  
GENERAL INFORMATION:

APPLICANT: Boone, Thomas  
APPLICANT: Li, Huimin

APPLICANT: Mann, Michael  
TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE

FILE REFERENCE: A-596  
CURRENT APPLICATION NUMBER: US/09/411,329C

CURRENT FILING DATE: 1999-10-01  
NUMBER OF SEQ ID NOS: 29

SOFTWARE: PatentIn version 3.0  
SEQ ID NO 26

LENGTH: 27  
TYPE: PRT

ORGANISM: Artificial  
FEATURE:

OTHER INFORMATION: Analog form of native pro-fibrolase of Agkistrodon contortrix

US-09-411-329C-26

Query Match 5.9%; Score 124; DB 4; Length 27;  
Best Local Similarity 87.5%; Pred. No. 1.9e-06;  
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 371 AYVEQKLISEDLNSAVDHHHHH 394  
Db 4 SFLEQKLISEDLNSAVDHHHHH 27



RESULT 10  
US-08-446-692-24  
; Sequence 24, Application US/08446692  
; Patent No. 5759551  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446.692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)415-8745  
; TELEFAX: (516)751-6849  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 37 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-446-692-24

Query Match 5.9%; Score 124; DB 1; Length 37;  
Best Local Similarity 84.0%; Pred. NO. 3.2e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 103 ALNIWDRFDVFCITGLASNGYIRGNS 127  
DB 1 ALNIWDRFDVFCITGLATGTYLKGNS 25  
RESULT 11  
US-08-488-351A-24  
; Sequence 24, Application US/08488351A  
; Patent No. 5843446  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488.351A  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/446.692  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/229.275  
FILING DATE: 14-APR-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/057.166  
FILING DATE: 27-APR-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria C.H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-488-351A-24

Query Match 5.9%; Score 124; DB 2; Length 37;  
Best Local Similarity 84.0%; Pred. NO. 3.2e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 103 ALNIWDRFDVFCITGLASNGYIRGNS 127  
DB 1 ALNIWDRFDVFCITGLATGTYLKGNS 25

RESULT 12  
US-09-433-428D-57  
; Sequence 57, Application US/09433428D  
; Patent No. 6149910  
; GENERAL INFORMATION:  
; APPLICANT: De Leys, Robert J.  
; APPLICANT: Zheng, Jian  
; TITLE OF INVENTION: PEPTIDES FOR THE DETECTION OF HIV-1 GROUP O  
; FILE REFERENCE: CDS-207  
; CURRENT APPLICATION NUMBER: US/09/433.428D  
; CURRENT FILING DATE: 1999-11-04  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 57  
; LENGTH: 439  
; TYPE: PPT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:HIV Consensus  
US-09-433-428D-57

Query Match 5.9%; Score 124; DB 4; Length 439;  
Best Local Similarity 57.8%; Pred. NO. 0.00017;  
Matches 26; Conservative 3; Mismatches 10; Indels 6; Gaps 1;  
QY 356 ERAAHVSGQFRFRYQAYV-----EOKLISEEDLNSAVDHHHHH 394  
||| ||| :| :| ||||| ||||| ||||| |||||



Db 395 ERDRDTSGRLVHGFLAIWVDLGPQKLISEEDLNSAVDHHHHH 439

## RESULT 13

US-09-433-428D-58  
; Sequence 58, Application US/09433428D  
; Patent No. 6149910  
; GENERAL INFORMATION:  
; APPLICANT: De Leys, Robert J.  
; APPLICANT: Zheng, Jian  
; TITLE OF INVENTION: PEPTIDES FOR THE DETECTION OF HIV-1 GROUP O  
; FILE REFERENCE: CDS-207  
; CURRENT APPLICATION NUMBER: US/09/433.428D  
; CURRENT FILING DATE: 1999-11-04  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: Patent In Release Ver. 2.0  
; SEQ ID NO 58  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:HIV Consensus  
; OTHER INFORMATION: Sequence  
US-09-433-428D-58

Query Match 5.8%; Score 121; DB 4; Length 368;  
Best Local Similarity 21.2%; Pred. No. 0.00027;  
Matches 82; Conservative 43; Mismatches 131; Indels 130; Gaps 16;  
QY 63 MGAKPTGSAANAATTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWD-----RF 110  
Db 59 MGRKWTWESPEKNRPLKDRINIVLSRELKEPP--RGAHFLAKSLDLDALRLIEQPELASKV 116  
QY 111 DVFCITGLASNGVIRNSTAFNLVGLGVKGTNNANELPNVSLNSGVVELYDTDFSWSV 170  
Db 117 DMVITVGSSVY---QEAHQPGHRLFLVTRI-----MQEFESDTFF----- 155  
QY 171 GARGALWECGATLGAEOYQAKSPKVELNVICNVSO-----FSYNKPKGVKGVAFP 223  
Db 156 -----PEIDLKYLKLLPEYPGVLSEVQEKIKYKFEYERKGRSARLL 200  
QY 224 LPTDAGVA-----TATGTKSATINHEWQVGASLSYRLNS----- 258  
Db 201 L---SGIVQQNNLLRATEAQHMLQLTAWGIKQLRARLQALETLMQNQRLNSWGCKGR 257  
QY 259 LVPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAW--NPSL---LGNATALSTDTDSFDF 314  
Db 258 ITCYTSARMH-----ASWSNKSLEDIWDNNMTNMQWQQQVNVV 294  
QY 315 MQIVSCQINKPKSKKACGVTGATLVADKW-SLTAEARLINERAHVSGOFREYQAYV 373  
Db 295 SSIIVDKILEAQDQOQENV---RELLEDKKWASLWNFDTN-----WLWYIKIFI 342  
QY 374 -----EQKLISEEDLNSAVDHHHHH 394  
Db 343 MIVGPEQKLISEEDLNSAVDHHHHH 368

## RESULT 14

US-08-519-385B-1  
; Sequence 1, Application US/08519385B  
; Patent No. 6001372  
; GENERAL INFORMATION:  
; APPLICANT: Demars, Robert I.  
; APPLICANT: Ortiz, Linette (n.m.i.)  
; TITLE OF INVENTION: Antigenic Peptides of Chlamydia trachomatis  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 411 East Wisconsin Avenue  
; CITY: Milwaukee  
; STATE: WI

COUNTRY: U.S.A.  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/519,385B  
FILING DATE:

CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Schwartz, Carl R.  
REGISTRATION NUMBER: 29,437  
REFERENCE/DOCKET NUMBER: 960296.93456  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5715  
TELEFAX: (414) 271-3552  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: no  
ANTI-SENSE: no  
FRAGMENT TYPE: internal fragment  
ORIGINAL SOURCE:  
ORGANISM: Chlamydia trachomatis  
STRAIN: Serovar E  
INDIVIDUAL ISOLATE: MOMP Protein  
US-08-519-385B-1

Query Match 5.7%; Score 119; DB 3; Length 31;  
Best Local Similarity 64.5%; Pred. No. 8e-06;  
Matches 20; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 240 TINHEWQVGASLSYRLNSLVPYIGVQWSRA 270  
Db 1 SIDHWEQASLSYRLNMFYIGVKWSRA 31

## RESULT 15

US-09-100-409A-51  
; Sequence 51, Application US/09100409A  
; Patent No. 6090388  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Chang Yi  
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR  
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND  
; TITLE OF INVENTION: IMMUNE DISORDERS  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
ZIP: 10154-0054  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version  
SOFTWARE: #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100,409A  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME:



```

;
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-100-409A-51

```

```

Query Match      5.5%; Score 115; DB 3; Length 25;
Best Local Similarity 80.0%; Pred. No. 1.5e-05;
Matches 20; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127
    |||||
DB 1 ALNIWDRFDVFTLGATGKLGNS 25

```

Search completed: February 7, 2002, 21:36:25  
Job time: 20494 sec







TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: APC  
US-08-452-654-7

Query Match 6.1%; Score 138; DB 1; Length 2842;  
Best Local Similarity 20.0%; Pred. No. 0.11;  
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERTPPADLSA-OGLEASANKSAEQAORTAGAEAPKE-----SKTDSV----- 55  
DB 1658 DLTIESPENELAAEGVGRGAQSGEFKRDPTPTGRSTDEAQQGKTSVTPPELDNKA 1717  
QY 56 ERWSTLRSVAVNALM-----SLADKLGIASSSSSTSRSDVDSTTATPTPP 103  
DB 1718 EEGDILAEICNSAMPKSHKPFVRVKIMDQVQASASSAPNKLNDGKKKPTSPVKP 1777  
QY 104 PPTSDYKTAQATYDT-----IFT-----STSLADIQAALVSLQDAV----- 141  
DB 1778 IPQNTYRTRVRKNADSKNNLNAERVFSDNKKQNLKNNKDFNDKLPNNEDVRGSGF 1837  
QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNQAAILDSL 191  
DB 1838 AFDSPHHYTPTEGTPYCFSRNDSLSLDFDDDDVLSREKAEL-----RKAKENK---ESE 1890  
QY 192 GKLTSDYKTAQATYDT-----IFT-----STSLADIQAALVSLQDAV----- 141  
DB 1891 AKVTS-----HTELTSNQOSANKTQAIKQPINRGPKPILO-KOSTFPQSSKDIPDRGAA 1945  
QY 248 QIEKDGNAIGDAYFAGQNASGAVENAKNSNISIDSKAAIAATAKTQIAEAKKFPDPS 307  
DB 1946 TDEKLON-----FAIENTPVCFSHNSLSLSDID----- 1975  
QY 308 ILQEAQWVIOAEKDLNKPAGSDVPNPGTTVGSKQ-----QGSSIGSIR 355  
DB 1976 --QENNN---KENEPKETEPPDSQGEPSKQASGYAPKSFHVEDTPVCFSRNLSLSLS 2030  
QY 356 VSMLLDDAENETASILMSGFRQIMHFNENTENPDSQAQOELAAQARAAGDSSAAAL 415  
DB 2031 I-----DSEDDLQECISS-----AMPKKKPSRLKGDNEKHSRPNMGILGED-LTLDL 2079  
QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQIASAAVVS 449  
DB 2080 KDIQRPDSEHGLSPDSENFWDKATQEGANSIVSSLHQAAAAACLS 2124

RESULT 11  
US-07-741-940-2  
Sequence 2, Application US/07741940  
Patent No. 5352775  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW

CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07741.940  
FILING DATE: 19920109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-741-940-2

Query Match 6.1%; Score 138; DB 1; Length 2842;  
Best Local Similarity 20.0%; Pred. No. 0.11;  
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;  
QY 11 DETERTPPADLSA-OGLEASANKSAEQAORTAGAEAPKE-----SKTDSV----- 55  
DB 1659 DLTIESPENELAAEGVGRGAQSGEFKRDPTPTGRSTDEAQQGKTSVTPPELDNKA 1718  
QY 56 ERWSTLRSVAVNALM-----SLADKLGIASSSSSTSRSDVDSTTATPTPP 103  
DB 1719 EEGDILAEICNSAMPKSHKPFVRVKIMDQVQASASSAPNKLNDGKKKPTSPVKP 1778  
QY 104 PPTSDYKTAQATYDT-----IFT-----STSLADIQAALVSLQDAV----- 141  
DB 1779 IPQNTYRTRVRKNADSKNNLNAERVFSDNKKQNLKNNKDFNDKLPNNEDVRGSGF 1838  
QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNQAAILDSL 191  
DB 1839 AFDSPHHYTPTEGTPYCFSRNDSLSLDFDDDDVLSREKAEL-----RKAKENK---ESE 1891  
QY 192 GKLTSDYKTAQATYDT-----IFT-----STSLADIQAALVSLQDAV----- 141  
DB 1892 AKVTS-----HTELTSNQOSANKTQAIKQPINRGPKPILO-KOSTFPQSSKDIPDRGAA 1946  
QY 248 QIEKDGNAIGDAYFAGQNASGAVENAKNSNISIDSKAAIAATAKTQIAEAKKFPDPS 307  
DB 1947 TDEKLON-----FAIENTPVCFSHNSLSLSDID----- 1976  
QY 308 ILQEAQWVIOAEKDLNKPAGSDVPNPGTTVGSKQ-----QGSSIGSIR 355  
DB 1977 --QENNN---KENEPKETEPPDSQGEPSKQASGYAPKSFHVEDTPVCFSRNLSLSLS 2031  
QY 356 VSMLLDDAENETASILMSGFRQIMHFNENTENPDSQAQOELAAQARAAGDSSAAAL 415  
DB 2032 I-----DSEDDLQECISS-----AMPKKKPSRLKGDNEKHSRPNMGILGED-LTLDL 2080  
QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQIASAAVVS 449  
DB 2081 KDIQRPDSEHGLSPDSENFWDKATQEGANSIVSSLHQAAAAACLS 2125  
RESULT 12  
US-08-289-548A-2  
Sequence 2, Application US/08289548A



```

; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-289-548A-2

Query Match 6.1%; Score 138; DB 1; Length 2843;
Best Local Similarity 20.0%; Pred. No. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERTPPADLSA-OGLEASAAKSAEQAQRIAGAEAKPKE-----SKTDSV----- 55
Db 1659 DLTIESPNEUAGGVGGAGGAGQSGFEKRDPTIPEGTSTDEAQGGKTSSTVPIELDDNKA 1718
QY 56 ERWSILSAVNALM-----SLADKLGITASSNSSSTSRSDVSTTATATPPP 103
Db 1719 EEGDILAEBCINSMPKGGKSHKPRVVKIMDQVQOASASSAPNKNQLDGKKKKPTSPVKP 1778
QY 104 PPTSDDKYKTAQYADT-----IFT-----STSLADIQAALVSLQDAV---- 141
Db 1779 IPQNTYEVTRVRKNWDSNNKLNNAERVFSDNKKSKQNLKNNKDFNDKLPNNEDRVRGSP 1838
QY 142 -----TNIKDT--AATDEETAIAAEWETKNAIDAKVGAQITELAKYASDNQAILDSL 191
Db 1839 AFDSPHHYTPTEGTFYCFYSRNDLSLDFDDDDVLSREKAEL---RKARENK---ESE 1891
QY 192 KGLTSFDLLQALLOSIVANNKKAELKEM-----QDNPPVPEKTPATAQSLVDQTDATAT 247
Db 1892 AKVTS-----HTELTSNQOASANKQIAKQPINRGQPKPILO-KOSTFPQSSKDIPDRGAA 1946
QY 248 QIEKDGNAGIDAYFAGQNASGAVENAKNSNISINIDSAKAIATAKTQIAEAQKFPDSP 307

```

```

Db 1947 TDEKLON-----FAIENTPVCFSHNSHSSLSLSDID----- 1976
QY 308 ILQEAQMVIAEKDLKNIKPADGSDVPNPGTTVGGSKQ-----QCSSISGISIR 355
Db 1977 --QENNN--KENEPKETEPDPSQGEPSKQASGYAPKSFHVEDTPVCFSRNSSL3SL5 2031
QY 356 VSMLLDDAENETASILMSGFROMIHMFNTENPDQAAQOELAAQAAKAAAGDDSAAL 415
Db 2032 I-----DSEDDLQECISS-----AMPKKKKPSRLKGDNEKHSPRNMGGILGED-LTLDL 2080
QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQIASAAVVS 449
Db 2081 KQIQRDPDSEHGLSPDSENFENFKAIQEGANSIVSLHQAAAAACLS 2125

RESULT 13
US-08-452-654-2
; Sequence 2, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-452-654-2

Query Match 6.1%; Score 138; DB 1; Length 2843;
Best Local Similarity 20.0%; Pred. No. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERTPPADLSA-OGLEASAAKSAEQAQRIAGAEAKPKE-----SKTDSV----- 55

```



Db 1659 DLTIESPNNELAAAGVGGAQSGEFKRDITPTGRSTDEAOGKTSSTVITPELDNKA 1718  
QY 56 ERWSILRSVAVNALM-----SLADKLGIASSNSSSTSRGADVDSTTATPTPP 103  
Db 1719 EEGDILAEICNSAMPKSKHKPFVRKIMDQVOQAASSAPNKNQLDGKKKPTSPVKP 1778  
QY 104 PPTSDDYKTAQATAYDT-----IFT-----STSLADIAALVSLQDAV--- 141  
Db 1779 IPQTEYTRVRKNADSKNNLNAERVSFNDKSKQKNNKSKDFNDKLPNNEDVRGSGF 1838  
QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVGAQITELAKYASDNOAILDSL 191  
Db 1839 AFDSPHHYPTIEGTPYCFSRNDSLSLDFDDDDVLSREKAEL---RKAKENK---ESE 1891  
QY 192 GKLTSDLLQTLQSVANNKKAELLKEM---QDNPVVPGKTPAQAQSLVDQTDATAT 247  
Db 1892 AKVTS-----HTELTSSNOQSANKTQAIKQPINRGQPKPILO-KOSTFPQSSKIDIPRGAA 1946  
QY 248 QIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAKAAIATAKTQIAEAQKKFPDPS 307  
Db 1947 TDEKLQN-----FAIENTPVCFSHNSLSLSDID----- 1976  
QY 308 ILQEAQOMVTOAEKDLKNKIPADGSDVPNPGTTVGGSKQ-----QGSIGSIR 355  
Db 1977 --QENNN---KENEPIKETEPDPSQGEPSKPAQSGYAPKSFHVEDTPVCFSRNSSLSSLS 2031  
QY 356 VSMLLDDAENETASILMSGFQMIHMENTENPDQQAQELAAQAAKAGDDSAAL 415  
Db 2032 I-----DSEDDLLOECISS-----AMPKKKPSRLKGDNEKHSRNMGGILGED-LTLDL 2080  
QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALQIAAAYVS 449  
Db 2081 KDIQRPDSEHGLSPDSENFQWKAIEGANSIVSLHQAAAAACLS 2125

## RESULT 14

US-08-452-655B-2  
; Sequence 2, Application US/08452655B  
; Patent No. 5783666  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
; NUMBER OF SEQUENCES: 102  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/452,655B  
; FILING DATE: 25-MAY-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,548  
; FILING DATE: 12-AUG-1994

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/741,940  
; FILING DATE: 08-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107.49964  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2843 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-452-655B-2

Query Match 6.1%; Score 138; DB 1; Length 2843;  
Best Local Similarity 20.0%; Pred. No. 0.11;  
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;  
QY 11 DETERPPADLSA-QGLEASAANKSAAQRIAGAEPKE-----SKTDSV----- 55  
Db 1659 DLTIESPNNELAAAGVGGAQSGEFKRDITPTGRSTDEAOGKTSSTVITPELDNKA 1718  
QY 56 ERWSILRSVAVNALM-----SLADKLGIASSNSSSTSRGADVDSTTATPTPP 103  
Db 1719 EEGDILAEICNSAMPKSKHKPFVRKIMDQVOQAASSAPNKNQLDGKKKPTSPVKP 1778  
QY 104 PPTSDDYKTAQATAYDT-----IFT-----STSLADIAALVSLQDAV--- 141  
Db 1779 IPQTEYTRVRKNADSKNNLNAERVSFNDKSKQKNNKSKDFNDKLPNNEDVRGSGF 1838  
QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVGAQITELAKYASDNOAILDSL 191  
Db 1839 AFDSPHHYPTIEGTPYCFSRNDSLSLDFDDDDVLSREKAEL---RKAKENK---ESE 1891  
QY 192 GKLTSDLLQTLQSVANNKKAELLKEM---QDNPVVPGKTPAQAQSLVDQTDATAT 247  
Db 1892 AKVTS-----HTELTSSNOQSANKTQAIKQPINRGQPKPILO-KOSTFPQSSKIDIPRGAA 1946  
QY 248 QIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAKAAIATAKTQIAEAQKKFPDPS 307  
Db 1947 TDEKLQN-----FAIENTPVCFSHNSLSLSDID----- 1976  
QY 308 ILQEAQOMVTOAEKDLKNKIPADGSDVPNPGTTVGGSKQ-----QGSIGSIR 355  
Db 1977 --QENNN---KENEPIKETEPDPSQGEPSKPAQSGYAPKSFHVEDTPVCFSRNSSLSSLS 2031  
QY 356 VSMLLDDAENETASILMSGFQMIHMENTENPDQQAQELAAQAAKAGDDSAAL 415  
Db 2032 I-----DSEDDLLOECISS-----AMPKKKPSRLKGDNEKHSRNMGGILGED-LTLDL 2080  
QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALQIAAAYVS 449  
Db 2081 KDIQRPDSEHGLSPDSENFQWKAIEGANSIVSLHQAAAAACLS 2125

## RESULT 15

US-08-452-655B-7  
; Sequence 7, Application US/08452655B  
; Patent No. 5783666  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.



APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,655B  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
US-08-452-655B-7

Query Match 6.1%; Score 138; DB 1; Length 2843;  
Best Local Similarity 20.0%; Pred. No. 0.11;  
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERPPADLSA-OGLEASAANKSAEAOIACAEAKPE-----SKTDSV----- 55  
DB 1659 DLTIESPPNELAAGEVGRGAQSGEKEKRTIETGRSTDEAOGGKTSVVTIPELDDNKA 1718  
QY 56 ERWSILRSVNAALM-----SLADKLGIASNSSTSRSDVDSTTATPTPP 103  
DB 1719 EEGDILAEICINSAMPKSHKPPRVKIMDQVOOASASSAPNKNQLDGKKKPTSPVKP 1778  
QY 104 PPTSDYKTKQAOTAYDT-----IFT-----STSLADIQAALVSLQDAV---- 141  
DB 1779 IPONTEYTRVRKNADSKNNLNAERVFSNKNDSKKNLNNKSKDFNDKLPNNEDVRVGRSF 1838  
QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDSL 191  
DB 1839 AFDSPPHYPIETGPTCFNSNDLSLDDDDVDLSREKAEI-----RKAKENK-----ESE 1891  
QY 192 GKLTSEDLQATALLQSVANNKAAELKEM-----QDNVVPVKTPAIAOSLYDQTDATAT 247  
DB 1892 AKVTS-----HTELTNQOSANKTQAIKQPINRGQPKPIQ-KQSTFPQSSKDIPIRGAA 1946  
QY 248 QIEKDGNAIGDAYFAGONAGVAKNSNNSISNIDSAKAAIATAKTQIAEAQKKFPDSP 307  
DB 1947 TDEKLQN-----FAIENTPVCFSHNSLSLSDID----- 1976

QY 308 ILQEAQMVIAQEKDLKNIKPADGSDVPNPNGTIVGSKQ-----QGSSTIGSIR 355  
DB 1977 --QENN---KENEPKETETPPDSOGEPKPOASGVAPKSEHVEDTPVCFSRNSSLSLS 2031  
QY 356 VSMLLDDAENETASIIIMSGFROMIHMFNTENPDSSQAQOELAAQARAAGDSDSAAAL 415  
DB 2032 I-----DSEDDLLQECISS-----AMPKKKPSRLKGDNEKHSRNMGGILGED-LTLDL 2080  
QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALQOIASAAVVS 449  
DB 2081 KDIQRPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAACLS 2125

Search completed: February 7, 2002, 21:36:24  
Job time: 20493 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:38:09 ; Search time 96.2 Seconds  
(without alignments)  
311.983 Million cell updates/sec

Title: US-09-391-606-15

Perfect score: 2103

Sequence: 1 MLPVGNPSPSLIDGTWE.....QKLISEEDLNSAYDHHHHH 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1947	92.6	389	2 A43587	major outer membra
2	1947	92.6	389	2 D86577	major outer membra
3	1910	90.8	389	2 I40864	major outer membra
4	1882	89.5	389	2 I40739	major outer membra
5	1534.5	73.0	389	1 MMCWP3	major outer membra
6	1496.5	71.2	389	2 A60109	major outer membra
7	1477	70.2	392	2 A40371	major outer membra
8	1419	67.5	402	2 I40740	major outer membra
9	1415	67.3	402	1 MMCWPM	major outer membra
10	1410	67.0	402	2 B60109	major outer membra
11	1408	67.0	402	2 A60341	major outer membra
12	1351.5	64.3	387	2 C81747	major outer membra
13	1351	64.2	404	2 I40741	major outer membra
14	1349.5	64.2	387	2 J70947	major outer membra
15	1349.5	64.2	387	2 S16034	mouse pneumonitis
16	1340.5	63.7	397	1 MMCWTH	major outer membra
17	1338.5	63.6	397	2 JF0413	major outer membra
18	1333.5	63.4	375	2 S11007	major outer membra
19	1328	63.1	374	2 S11006	major outer membra
20	1325	63.0	396	2 S12799	major outer membra
21	1315.5	62.6	393	1 MMCWTE	major outer membra
22	1315.5	62.6	397	1 MMCWTC	major outer membra
23	1313.5	62.5	393	2 S06259	major outer membra
24	1310.5	62.3	393	2 T01645	major outer membra
25	1308.5	62.2	393	2 H71484	probable major out
26	1307.5	62.2	395	1 MMCWTF	major outer membra
27	1303.5	62.0	393	2 JC1432	major outer membra
28	1295	61.6	394	2 S11012	major outer membra
29	1292	61.4	372	2 B60756	major outer membra

ALIGNMENTS

RESULT 1

A43587

major outer membrane protein, porin CP0051 precursor [imported] - Chlamydomophila pneu  
N:Alternate names: MOMP  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 11-May-2000  
C:Accession: A43587; A49751; A49216; G72044; F81619  
R:Perez Melçosa, M.; Kuo, C.C.; Campbell, L.A.  
Infect. Immun. 59, 2195-2199, 1991  
A:Title: Sequence analysis of the major outer membrane protein gene of Chlamydia pne  
A:Reference number: A43587; MUID:91244474  
A:Accession: A43587  
A:Molecule type: DNA  
A:Residues: 1-389 <PCR>  
A:Cross-references: GB:M69230; NID:g144540; PIDN:AAA73071.1; PID:g144541  
R:Cartier, M.W.; Al-Mahdawi, S.A.H.; Giles, I.G.; Treharne, J.D.; Ward, M.E.; Clarke,  
J. Gen. Microbiol. 137, 465-475, 1991  
A:Title: Nucleotide sequence and taxonomic value of the major outer membrane protein  
A:Reference number: A49751; MUID:91237311  
A:Accession: A49751  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-389 <CAR>  
A:Cross-references: GB:M64064; GB:M34942; NID:g144534; PIDN:AAA23143.1; PID:g144535  
A:Note: isolate IOL-207  
R:Gaydos, C.A.; Quinn, T.C.; Bobo, L.D.; Eiden, J.J.  
Infect. Immun. 60, 5319-5323, 1992  
A:Title: Similarity of Chlamydia pneumoniae strains in the variable domain IV region  
A:Reference number: A49216; MUID:93084388  
A:Accession: A49216  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1297-352 <GAY>  
A:Cross-references: GB:S50607; NID:g260972; PIDN:AAB24363.1; PID:g260973  
A:Note: sequence extracted from NCBI backbone (NCBIN:120604, NCBIPI:120605)  
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: G72044  
A:Molecule type: DNA  
A:Residues: 1-389 <ARN>  
A:Cross-references: GB:AF001652; GB:AF001363; NID:g4376997; PIDN:AA018834.1; PID:g43  
A:Experimental source: strain CWL029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hick  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR3  
A:Reference number: A81500; MUID:20150255  
A:Accession: F81619  
A>Status: preliminary  
A:Molecule type: DNA

30 1285 61.1 372 2 S11009 major outer membra  
31 1275 60.6 394 1 MMCWTB major outer membra  
32 573 17.7 158 2 A60333 outer membrane pro  
33 561 17.2 156 2 B60333 outer membrane pro  
34 355.5 16.9 155 2 C60333 outer membrane pro  
35 215.5 10.2 136 2 F30593 outer membrane pro  
36 215 10.2 135 2 C30593 outer membrane pro  
37 212.5 10.1 136 2 D30593 outer membrane pro  
38 206.5 9.9 136 2 B30593 outer membrane pro  
39 206.5 9.8 136 2 I30587 outer membrane pro  
40 206.5 9.8 136 2 E30593 outer membrane pro  
41 203.5 9.7 136 2 A30593 outer membrane pro  
42 189 9.0 340 2 H71479 probable outer mem  
43 188.5 9.0 134 2 G30587 outer membrane pro  
44 182 8.7 340 2 H81742 major outer membra  
45 179.5 8.5 134 2 H30587 outer membrane pro



A:Residues: 1-389 <REA>  
A:Cross-references: GB:AB002168; GB:AB002161; NID:g7188982; PIDN:AAF37944.1; PID:g718899  
A:Experimental source: strain AR39, HL cells  
C:Genetics:

A:Gene: ompA; CP0051  
C:Superfamily: Chlamydia major outer membrane protein  
C:Keywords: membrane protein  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 92.6%; Score 1947; DB 2; Length 389;  
Best Local Similarity 100.0%; Pred. No. 2.7e-148;  
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDASISLRAGFYGDYVDFRILKVDAPKTF 61  
DB 24 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDASISLRAGFYGDYVDFRILKVDAPKTF 83  
QY 62 SMGAKPTGSAANVTAVDRENPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 121  
DB 84 SMGAKPTGSAANVTAVDRENPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 143  
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 181  
DB 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 203  
QY 182 ATLGAEFQYAOQSKPKVEELNVICNVSOFSVKNPKYKGVAFPLPTDAGVATATGKTSATI 241  
DB 204 ATLGAEFQYAOQSKPKVEELNVICNVSOFSVKNPKYKGVAFPLPTDAGVATATGKTSATI 263  
QY 242 NYHEWQVGSLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301  
DB 264 NYHEWQVGSLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323  
QY 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 361  
DB 324 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 383  
QY 362 SQQFRF 367  
DB 384 SQQFRF 389

RESULT 2

D86577  
major outer membrane protein [imported] - Chlamydia pneumoniae (strain J138)  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: D86577  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: D86577  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-389 <STC>  
A:Cross-references: GB:BA000008; NID:g8979067; PIDN:BA989802.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: ompA  
C:Superfamily: Chlamydia major outer membrane protein

Query Match 92.6%; Score 1947; DB 2; Length 389;  
Best Local Similarity 100.0%; Pred. No. 2.7e-148;  
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDASISLRAGFYGDYVDFRILKVDAPKTF 61  
DB 24 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDASISLRAGFYGDYVDFRILKVDAPKTF 83

QY 62 SMGAKPTGSAANVTAVDRENPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 121  
DB 84 SMGAKPTGSAANVTAVDRENPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 143  
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 181  
DB 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 203  
QY 182 ATLGAEFQYAOQSKPKVEELNVICNVSOFSVKNPKYKGVAFPLPTDAGVATATGKTSATI 241  
DB 204 ATLGAEFQYAOQSKPKVEELNVICNVSOFSVKNPKYKGVAFPLPTDAGVATATGKTSATI 263  
QY 242 NYHEWQVGSLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301  
DB 264 NYHEWQVGSLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323  
QY 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 361  
DB 324 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 383  
QY 362 SQQFRF 367  
DB 384 SQQFRF 389

RESULT 3

I40864  
major outer membrane protein - Chlamydia psittaci  
C:Species: Chlamydia psittaci, Chlamydia psittaci  
C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 31-Mar-2000  
C:Accession: I40864; S33465  
R:Gijses, A.A.; Carrick, F.N.; Lavin, M.F.  
Gene 138, 139-142, 1994  
A:Title: Remarkable sequence relatedness in the DNA encoding the major outer membra  
A:Reference number: I40864; MUID:94171025  
A:Accession: I40864  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-389 <RES>  
A:Cross-references: EMBL:X72023; NID:g313844; PIDN:CAA50906.1; PID:g313845  
C:Superfamily: Chlamydia major outer membrane protein

Query Match 90.8%; Score 1910; DB 2; Length 389;  
Best Local Similarity 97.5%; Pred. No. 2.5e-145;  
Matches 357; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDASISLRAGFYGDYVDFRILKVDAPKTF 61  
DB 24 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDASISLRAGFYGDYVDFRILKVDAPKTF 83  
QY 62 SMGAKPTGSAANVTAVDRENPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 121  
DB 84 SMGAKPTGSAANVTAVDRENPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 143  
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 181  
DB 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 203  
QY 182 ATLGAEFQYAOQSKPKVEELNVICNVSOFSVKNPKYKGVAFPLPTDAGVATATGKTSATI 241  
DB 204 ATLGAEFQYAOQSKPKVEELNVICNVSOFSVKNPKYKGVAFPLPTDAGVATATGKTSATI 263  
QY 242 NYHEWQVGSLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301  
DB 264 NYHEWQVGSLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323  
QY 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 361  
DB 324 TTTLATSDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 383  
QY 362 SQQFRF 367  
DB 384 SQQFRF 389



Db 384 SQQFRF 389

# RESULT 4

major outer membrane protein precursor - Chlamydomonadales (strain equine/N16)

C:Species: Chlamydomonadales, Chlamydomonadales

A:Variety: strain equine/N16

C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 20-Apr-2000

C:Accession: I40739

R:Storey, C.; Lusher, M.; Yates, P.; Richmond, S.

J. Gen. Microbiol. 139, 2621-2626, 1993

A:Title: Evidence for Chlamydomonadales of non-human origin.

A:Reference number: I40739; MUID:94103736

A:Accession: I40739

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-389 <STO>

A:Cross-references: GB:L04982; NID:g289840; PIDN:AA17397.1; PID:g289841

C:Comment: On the basis of the major outer membrane protein the authors classified the

the sequence of the genome strain CWL029 and strain IOL-207. See PIR:A43587.

C:Genetics:

A:Gene: momp

C:Superfamily: Chlamydia major outer membrane protein

C:Keywords: membrane protein

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 89.5%; Score 1882; DB 2; Length 389;

Best Local Similarity 94.5%; Pred. No. 4.3e-143;

Matches 346; Conservative 15; Mismatches 5; Indels 0; Gaps 0;

QY 2 LPVGNPSPDLLIDGTIWEAGAAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61

DB 24 LPVGNPSPDLLIDGTIWEAGAAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 83

QY 62 SMGAKPTGSAANYTTAVDRPNPAYNKLHDAEFTNAGFIATLNIWDRFDFVCTLGASNG 121

DB 84 SMGAKPTGSAANYTTAVDRPNPAYNKLHDAEFTNAGFIATLNIWDRFDFVCTLGASNG 143

QY 122 YIRGNSTAFNLVGLFGVKGTTVNAELPNVLSNGVVELYDTDSFSWSVGARGALWECG 181

DB 144 YKGNAAFNVLVGLFGVKGTTVNAELPNVLSNGVIELYDTTTFWSVGARGALWECG 203

QY 182 ATLGAEEFYAQSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGTSKATI 241

DB 204 ATLGAEEFYAQSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGTSKATI 263

QY 242 NYHEWQVGLSLYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGN 301

DB 264 NYHEWQVGLSLYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGN 323

QY 302 ATALSTTDSFDFMIVSCQINKEFKSRKACGVTGATLVADKWSLTAEARLINERAHHV 361

DB 324 ATAVSSDDQSFDFMIVSCQINKEFKSRKACGVTGATLVADKWSLTAEARLINERAHHV 383

QY 362 SQQFRF 367

DB 384 SQQFRF 389

# RESULT 5

MMCW3

major outer membrane protein precursor - Chlamydomonadales (strain S26/3)

C:Species: Chlamydomonadales, Chlamydomonadales

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 31-Mar-2000

C:Accession: S08770

R:Herring, A.J.; Tan, T.W.; Baxter, S.; Inglis, N.F.; Dunbar, S.

FEMS Microbiol. Lett. 65, 153-158, 1989

A:Title: Sequence analysis of the major outer membrane protein gene of an ovine abortion

A:Reference number: S08770

A:Accession: S08770

Query Match 71.2%; Score 1496.5; DB 2; Length 389;

Best Local Similarity 73.9%; Pred. No. 3.2e-112;

Matches 272; Conservative 42; Mismatches 51; Indels 3; Gaps 2;

QY 2 LPVGNPSPDLLIDGTIWEAGAAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61

DB 23 LPVGNPSPDLLIDGTIWEAGAAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 82

QY 62 SMGAKPTGSAANYTTAVDRPNPAYNKLHDAEFTNAGFIATLNIWDRFDFVCTLGASNG 121

DB 83 SMGAKPTGSAANYTTAVDRPNPAYNKLHDAEFTNAGFIATLNIWDRFDFVCTLGASNG 142

A:Molecule type: DNA

A:Residues: 1-389 <HER>

A:Cross-references: EMBL:X51859; NID:g40600; PIDN:CAA36152.1; PID:g40601

C:Superfamily: Chlamydia major outer membrane protein

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 73.0%; Score 1534.5; DB 1; Length 389;

Best Local Similarity 75.7%; Pred. No. 2.9e-115;

Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;

QY 2 LPVGNPSPDLLIDGTIWEAGAAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61

DB 23 LPVGNPSPDLLIDGTIWEAGAAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 82

QY 62 S-MGAKPTGSAANYTTAVDRPNPAYNKLHDAEFTNAGFIATLNIWDRFDFVCTLGASNG 120

DB 83 TMGAVPTGTAAANYKTPTDRPNIAIYKHLQDAEFTNAGFIATLNIWDRFDFVCTLGASNG 142

QY 121 GYIRGNSTAFNLVGLFGVKGTTVNAELPNVLSNGVVELYDTDSFSWSVGARGALWECG 180

DB 143 GYIRGNSTAFNLVGLFGVKGTTVNAELPNVLSNGVVELYDTDSFSWSVGARGALWECG 202

QY 181 CATLGAEEFYAQSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGTSKAT 240

DB 203 CATLGAEEFYAQSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGTSKAT 262

QY 241 INYHEWQVGLSLYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLG 300

DB 263 IKYHEWQVGLSLYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLG 322

QY 301 NATALSTTDSFDFMIVSCQINKEFKSRKACGVTGATLVADKWSLTAEARLINERAHH 360

DB 323 NATALSTTDSFDFMIVSCQINKEFKSRKACGVTGATLVADKWSLTAEARLINERAHH 382

QY 361 VSGQFRF 367

DB 383 MNAQFRF 389

# RESULT 6

A60109

major outer membrane protein precursor - Chlamydomonadales (strain Guinea pig 1

C:Species: Chlamydomonadales, Chlamydomonadales

C:Date: 10-Nov-1992 #sequence\_revision 10-Nov-1992 #text\_change 31-Mar-2000

C:Accession: A60109

R:Zhang, Y.X.; Morrison, S.G.; Caldwell, H.D.; Baehr, W.

Infect. Immun. 57, 1621-1625, 1989

A:Title: Cloning and sequence analysis of the major outer membrane protein genes of

A:Reference number: A60109; MUID:89212917

A:Accession: A60109

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-389 <ZHA>

C:Superfamily: Chlamydia major outer membrane protein

C:Keywords: membrane protein

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 71.2%; Score 1496.5; DB 2; Length 389;

Best Local Similarity 73.9%; Pred. No. 3.2e-112;

Matches 272; Conservative 42; Mismatches 51; Indels 3; Gaps 2;

QY 2 LPVGNPSPDLLIDGTIWEAGAAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61

DB 23 LPVGNPSPDLLIDGTIWEAGAAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 82

QY 62 SMGAKPTGSAANYTTAVDRPNPAYNKLHDAEFTNAGFIATLNIWDRFDFVCTLGASNG 121

DB 83 SMGAKPTGSAANYTTAVDRPNPAYNKLHDAEFTNAGFIATLNIWDRFDFVCTLGASNG 142







F.1-22/Domain: signal sequence #status predicted <SIG>  
F.23-402/Product: major outer membrane protein #status predicted <MAT>

Query Match 67.3%; Score 1415; DB 1; Length 402;  
Best Local Similarity 70.3%; Pred. No. 1.1e-105;  
Matches 267; Conservative 37; Mismatches 62; Indels 14; Gaps 6;

QY 2 LPVGNPSPSLIDGTWEGAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61  
DB 23 LPVGNPAEPLSLIDGTWEGASGDCPCATWCDALSLRAGFYGYVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSAA-ANYTTAVDRPNPAYNKHLEDAEFTNAGFIALNIWDRFDFVCTL 116  
DB 83 SCMAATPTQATGNASNTQPEANGRPNTAYGRHMDAEWFSNAEFLALNIWDRFDFCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG---LPNVLSNGVVELYTDTSFWSVGA 172  
DB 143 GASNGYFKSSAAFNVLVGLIGFSSTSTSLPMLPNVGIQTGVVEFTDTSFWSVGA 202

QY 173 RGALWECGCATLGAEOAQSQPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 230  
DB 203 RGALWECGCATLGAEOAQSQPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 262

QY 231 ATATGKTSATINHEWQVQASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290  
DB 263 TEATDTKSAITIKYHEWQVGLALSRLNMLVPYIGVWNSRATFDADTIRIAQPKLSEILN 322

QY 291 LTAWNPSLLGNATALSTT---DSFSDFMQIVSCQINKKFSKACGVTGATLVADAKWSL 347  
DB 323 ITTWNPSLLGSTTTLPNNGCKDVLSDVLOIASIQINKMKSRKACGVAVGATLIDAKWSI 382

QY 348 TAEARLINERAHVSCQFRF 367  
DB 383 TGEARLINERAHVSCQFRF 402

RESULT 10  
B60109  
major outer membrane protein precursor - Chlamydomophila psittaci (strain meningopneumoni  
C:Species: Chlamydomophila psittaci, Chlamydia psittaci  
C:Date: 10-Nov-1992 #sequence\_revision 10-Nov-1992 #text\_change 31-Mar-2000  
C:Accession: B60109  
R:Zhang, Y.X.; Morrison, S.G.; Caldwell, H.D.; Baehr, W.  
Infect. Immun. 57, 1621-1625, 1989  
A:Title: Cloning and sequence analysis of the major outer membrane protein genes of two  
A:Reference number: A60109; MUID:89212917  
A:Accession: B60109  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-402 <ZHA>  
C:Superfamily: Chlamydia major outer membrane protein  
C:Keywords: membrane protein  
F.1-22/Domain: signal sequence #status predicted <SIG>  
F.23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 67.0%; Score 1410; DB 2; Length 402;  
Best Local Similarity 70.3%; Pred. No. 2.8e-105;  
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSPSLIDGTWEGAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61  
DB 23 LPVGNPAEPLSLIDGTWEGASGDCPCATWCDALSLRAGFYGYVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSAA-ANYTTAVDRPNPAYNKHLEDAEFTNAGFIALNIWDRFDFVCTL 116  
DB 83 SCMAATPTQATGNASNTQPEANGRPNTAYGRHMDAEWFSNAEFLALNIWDRFDFCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG---TTVNANELPNVLSNGVVELYTDTSFWSVGA 172  
DB 143 GASNGYFKSSAAFNVLVGLIGFSSTSTSLPMLPNVGIQTGVVEFTDTSFWSVGA 202

QY 173 RGALWECGCATLGAEOAQSQPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 230  
DB 203 RGALWECGCATLGAEOAQSQPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 262

QY 231 ATATGKTSATINHEWQVQASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290  
DB 263 TEATDTKSAITIKYHEWQVGLALSRLNMLVPYIGVWNSRATFDADTIRIAQPKLSEILN 322

QY 291 LTAWNPSLLGNATAL---STTDSFSDFMQIVSCQINKKFSKACGVTGATLVADAKWSL 347  
DB 323 ITTWNPSLLGSTTTLPNNGCKDVLSDVLOIASIQINKMKSRKACGVAVGATLIDAKWSI 382

QY 348 TAEARLINERAHVSCQFRF 367  
DB 383 TGEARLINERAHVSCQFRF 402

RESULT 11  
A60341  
major outer membrane protein precursor - Chlamydomophila psittaci (strain 6BC)  
C:Species: Chlamydomophila psittaci, Chlamydia psittaci  
C:Date: 11-Dec-1992 #sequence\_revision 24-Feb-1994 #text\_change 31-Mar-2000  
C:Accession: A44565; A60341; B60341  
R:Everett, K.D.E.  
submitted to the EMBL Data Library, December 1990  
A:Reference number: A44565  
A:Accession: A44565  
A:Molecule type: DNA  
A:Residues: 1-402 <EVE>  
A:Cross-references: GB:X56980; NID:g40568; PIDN:CAA40300.1; PID:g40569  
R:Everett, K.D.E.; Andersen, A.A.; Plaunt, M.; Hatch, T.P.  
Infect. Immun. 59, 2853-2855, 1991  
A:Title: Cloning and sequence analysis of the major outer membrane protein gene of  
A:Reference number: A60341; MUID:91310346  
A:Accession: A60341  
A:Molecule type: protein  
A:Residues: 23-35 <EV2>  
A:Accession: B60341  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 112-232; 317-349 <EV3>  
A:Cross-references: GB:X56980  
C:Superfamily: Chlamydia major outer membrane protein  
C:Keywords: membrane protein  
F.1-22/Domain: signal sequence #status predicted <SIG>  
F.23-402/Product: major outer membrane protein #status experimental <MAT>

Query Match 67.0%; Score 1408; DB 2; Length 402;  
Best Local Similarity 70.3%; Pred. No. 4e-105;  
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSPSLIDGTWEGAGDPCDPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61  
DB 23 LPVGNPAEPLSLIDGTWEGASGDCPCATWCDALSLRAGFYGYVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSAA-ANYTTAVDRPNPAYNKHLEDAEFTNAGFIALNIWDRFDFVCTL 116  
DB 83 SCMAATPTQATGNASNTQPEANGRPNTAYGRHMDAEWFSNAEFLALNIWDRFDFCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG---TTVNANELPNVLSNGVVELYTDTSFWSVGA 172  
DB 143 GASNGYFKSSAAFNVLVGLIGFSSTSTSLPMLPNVGIQTGVVEFTDTSFWSVGA 202

QY 173 RGALWECGCATLGAEOAQSQPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 230  
DB 203 RGALWECGCATLGAEOAQSQPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 262

QY 231 ATATGKTSATINHEWQVQASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290  
DB 263 TEATDTKSAITIKYHEWQVGLALSRLNMLVPYIGVWNSRATFDADTIRIAQPKLSEILN 322

QY 291 LTAWNPSLLGNATAL---STTDSFSDFMQIVSCQINKKFSKACGVTGATLVADAKWSL 347



Db 323 ITTWPSIGSTALPNNNGKDVLSVLIQIASIQINKMKSRKACGVAVGATLVDADKWSI 382  
QY 348 TAEARLINERAHVSGQFRF 367  
Db 383 TGEARLINERAHVNAQFRF 402

## RESULT 12

C81747

major outer membrane protein, porin TC0052 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000

C:Accession: C81747

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidberg, J.F.; White, O.; Hickey,

C.C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: AB1500; MUID:20150255

A:Accession: C81747

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-387 <TEP>

A:Cross-references: GB:AE002272; GB:AE002160; NID:G7190083; PIDN:AAF38941.1; PID:G719009

A:Experimental source: strain Nigg (MoPn)

C:Genetics:

A:Gene: TC0052

C:Superfamily: Chlamydia major outer membrane protein

Query Match 64.3%; Score 1351.5; DB 2; Length 387;

Best Local Similarity 66.4%; Pred. No. 1.3e-100;

Matches 245; Conservative 50; Mismatches 67; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCPCATWCDAISLRAGFYGDYVDFRILKYDAPKTF 61  
Db 23 LPVGNPAEPLMIDGILWEGFGDPCDCTTWCDAISLRGLGYGDFVDRVLKTDVNVKQF 82

QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKKHLHDAEFTNAGFIALNIWDRFVFCITLGA 118  
Db 83 ENGAAPTGD--ADLTATTPASRENPAYGKHQDAEMFTNAAYMALNIWDRFVFCITLGA 140

QY 119 SNGYIRGNSTAFNLVGLFGVGTNNANLPLNVSLSNGVVELYTDTSFSWSVGARGALWE 178  
Db 141 TSGYLKGNASAFNLVGLFGRDETAAVADDIPNVLSQAVVELYTDTAFAWSVGARGALWE 200

QY 179 CCATLGAFFOYAKSPKVEELNVCNVSQFVSNKPKYKGVAFPLPTDAGVATATGKS 238  
Db 201 CCATLGAFFOYAKSPKVEELNVCNVAEFTINPKGYGVEFFPLNPKAGTVSATDTKD 260

QY 239 ATINHEWQVGSASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 298  
Db 261 ASIDYHEWQASLALSYRLNMFYIGVKSRSASFADTIRIAQPKLPTAVLNLTAWNPTI 320

QY 299 LGNATALSTDSFDFMQIVSCQINKFKSRACGVTVGATLVADKWSLTAEARLINERA 358  
Db 321 SSGSGIDVDT--KITDTLQIVSLQNLNKKSRKSCGLAIGTITVDKADYAVTVETRLIDERA 378

QY 359 AHVSGQFRF 367  
Db 379 AHVNAQFRF 387

## RESULT 13

I40741

major outer membrane protein - Chlamydia trachomatis

C:Species: Chlamydia trachomatis

C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 20-Aug-1999

C:Accession: I40741

R:Zhang, Y.X.; Fox, J.G.; Ho, Y.; Zhang, L.; Stills, H.F.; Smith, T.F.

Mol. Biol. Evol. 10, 1327-1342, 1993

A:Title: Comparison of the major outer-membrane protein (MOMP) gene of mouse pneumonitis

A:Reference number: A49379; MUID:94104488

A:Accession: I40741

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-404 <RES>

A:Cross-references: GB:LI9221; NID:9410146; PIDN:AAA16615.1; PID:9410147

C:Genetics:

A:Gene: momp

C:Superfamily: Chlamydia major outer membrane protein

Query Match 64.2%; Score 1351; DB 2; Length 404;

Best Local Similarity 65.0%; Pred. No. 1.5e-100;

Matches 249; Conservative 50; Mismatches 66; Indels 18; Gaps 4;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCPCATWCDAISLRAGFYGDYVDFRILKYDAPKTF 61  
Db 23 LPVGNPAEPLMIDGILWEGFGDPCDCTTWCDAISLRGLGYGDFVDRVLKTDVNVKQF 82

QY 62 SMGAKPTGS---AAANYTTAVDRPNPAYNKKHLHDAEFTNAGFIALNIWDRFVFCITLGA 118  
Db 83 EMGPVPTTDTDAADITSTPRENPAYGKHQDAEMFTNAAYMALNIWDRFVFCITLGA 142

QY 119 SNGYIRGNSTAFNLVGLFGVGTNNANLPLNVSLSNGVVELYTDTSFSWSV 170  
Db 143 TSGYLKGNASAFNLVGLFG-DGVANAANAIAATVAADSLPNVLSQAVVELYTDTAFAWSV 201

QY 171 GARGALWECGCATLGAFFOYAKSPKVEELNVCNVSQFVSNKPKYKGVAFPLPTDAGV 230  
Db 202 GARGALWECGCATLGAFFOYAKSPKVEELNVCNVAEFTINPKGYGVEFFPLALTAGT 261

QY 231 ATATGTSKATINHYHEWQVGSASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290  
Db 262 DSATDTKDASIDYHEWQASLALSYRLNMFYIGVKSRSASFADTIRIAQPKLALAILD 321

QY 291 LTANPSSLGNAT-----ALSTTDSFDFMQIVSCQINKFKSRACGVTVGATLVADK 344  
Db 322 VTTNPTTIAGAGTTADGTGAAATANGLATLQIVSLQNLNKKSRKSCGLAIGTITVDADK 381

QY 345 WSLTAEARLINERAHVSGQFRF 367  
Db 382 YAVTVETRLIDERAHVNAQFRF 404

## RESULT 14

JT0947

mouse pneumonitis major outer membrane protein - Chlamydia trachomatis

C:Species: Chlamydia trachomatis

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Aug-1999

C:Accession: JT0947

R:Fielder, T.J.; Pal, S.; Peterson, E.M.; de la Maza, L.M.

A:Title: Sequence of the gene encoding the major outer membrane protein of the mouse

Gene 106, 137-138, 1991

A:Reference number: JT0947; MUID:92039057

A:Accession: JT0947

A:Molecule type: DNA

A:Residues: 1-387 <FIE>

A:Cross-references: GB:M64171; NID:g144536; PIDN:AAA23144.1; PID:g144537

C:Comment: C. trachomatis are Gram negative bacteria that cause a variety of infecti

C:Comment: This protein is strongly antigenic and elicits neutralizing antibodies an

C:Genetics:

A:Gene: om1MOPn

C:Superfamily: Chlamydia major outer membrane protein

C:Keywords: membrane protein

Query Match 64.2%; Score 1349.5; DB 2; Length 387;

Best Local Similarity 66.1%; Pred. No. 1.9e-100;

Matches 244; Conservative 50; Mismatches 68; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCPCATWCDAISLRAGFYGDYVDFRILKYDAPKTF 61  
Db 23 LPVGNPAEPLMIDGILWEGFGDPCDCTTWCDAISLRGLGYGDFVDRVLKTDVNVKQF 82



QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFDVFTLGA 118  
 DB 83 EMGAAPTGD--ADLTAPTASRENPAKGMQDAEMFTNAYMALNIWDRFDVFTLGA 140  
 QY 119 SNGYIRGNSTANLVGLFGVKTNNANLNPVLSNGVVELYTDTSFWSVGCARGALWE 178  
 DB 141 TSGYLKGNAAFNVLGFRDETAADIPNVLSQAQVVELYTDTAFAWSVGARAFAWE 200  
 QY 179 CGCATLGAEFOYQAQSKPKVEELNVCNVSOFSVKNPKGYKGVAFPLPTDAGVATATGKS 238  
 DB 201 CGCPTLGASFQQAQSKPKVEELNVCNAAEFTINPKGYGQEFPLNINAGTVSATDTKD 260  
 QY 239 ATINHEWQVGSLSYRLNSLPYIGVQWRSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 298  
 DB 261 ASIDTHEWQASLSYRLNMFTPYIGVKWSRASFDADTIRIAQPKLETSILKMTTWNPTI 320  
 QY 299 LGNATALSTTDSFSDFMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERA 358  
 DB 321 SSGSIDVDVT--KITDTLQIVSLQNLNKMKSRCGLAIGTTIVDADKYAVTVETRLIDERA 378  
 QY 359 AHVSGQFRF 367  
 DB 379 AHVNAQFRF 387

RESULT 15

SI6034  
 major outer membrane protein -- Chlamydia trachomatis  
 C:Species: Chlamydia trachomatis  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 12-Apr-1996  
 C:Accession: SI6034; A43875  
 R:Zhang, Y.X.; Zhang, L.  
 submitted to the EMBL Data Library, July 1991  
 A:Description: The nucleotide sequence of major outer membrane protein gene of mouse bio  
 A:Reference number: SI6034  
 A:Accession: SI6034  
 A>Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-387 <ZHA>  
 A:Cross-references: EMBL:X60678  
 R:Peterson, E.M.; Cheng, X.; Markoff, B.A.; Fielder, T.J.; de la Maza, L.M.  
 Infect. Immun. 59, 4147-4153, 1991  
 A:Title: Functional and structural mapping of Chlamydia trachomatis species-specific ma  
 A:Reference number: A43875; MUID:92040090  
 A:Accession: A43875  
 A>Status: Preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Experimental source: mouse pneumonitis strain  
 A:Note: sequence extracted from NCBI backbone (NCBIP:62877)  
 C:Superfamily: Chlamydia major outer membrane protein

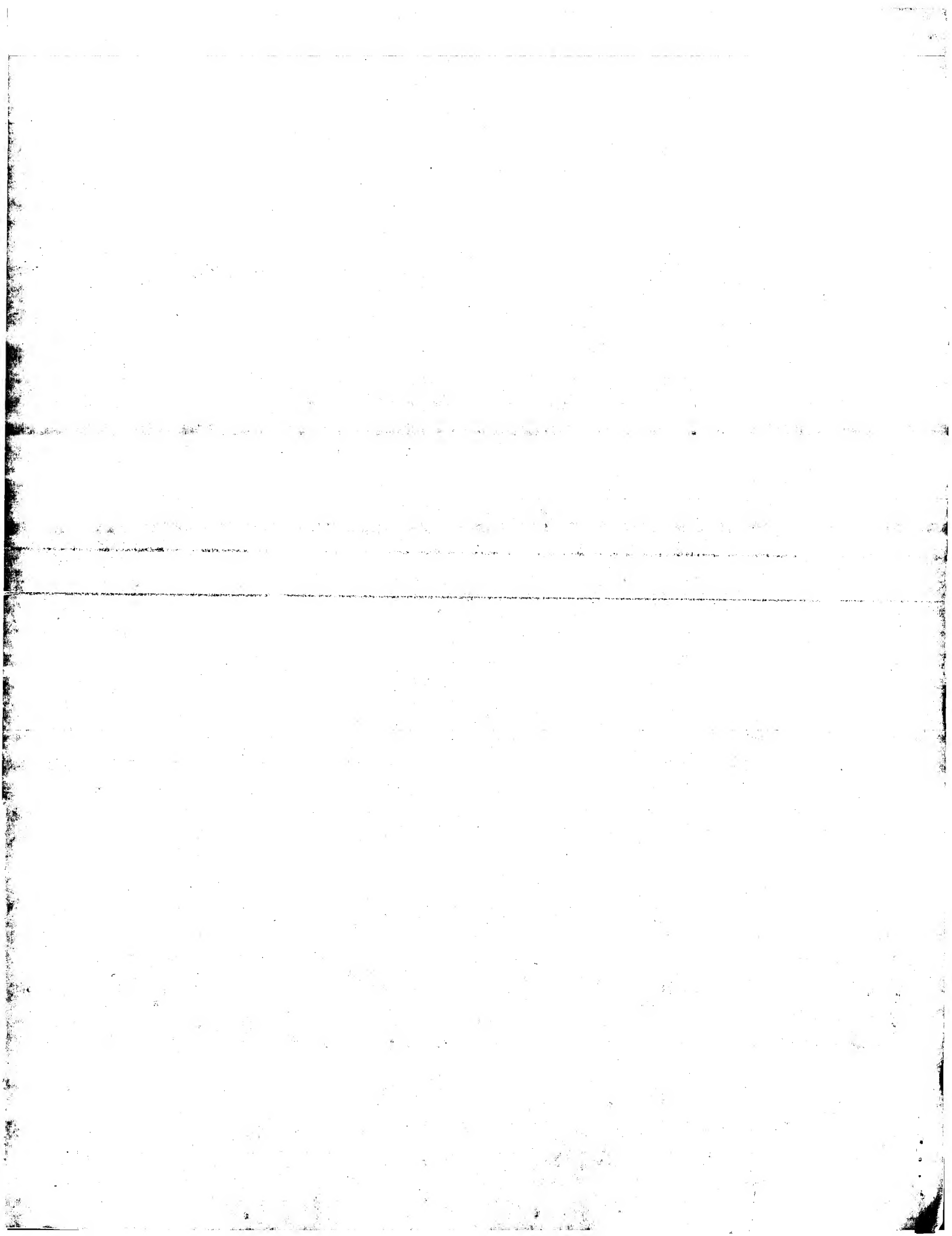
Query Match 64.2%; Score 1349.5; DB 2; Length 387;  
 Best Local Similarity 66.1%; Pred. No.1.9e-100;  
 Matches 244; Conservative 50; Mismatches 68; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLIDGTINWGAAGDPCDCAITWCDALSRAGFYGDYVDFRILKVDAPKTF 61  
 DB 23 LPVGNPAEPLMIDGILWEGFGDPCDCTTWCDAISLRGLYGGDFVDFRVLKTDVKNQF 82  
 QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFDVFTLGA 118  
 DB 83 EMGAAPTGD--ADLTAPTASRENPAKGMQDAEMFTNAYMALNIWDRFDVFTLGA 140  
 QY 119 SNGYIRGNSTANLVGLFGVKTNNANLNPVLSNGVVELYTDTSFWSVGCARGALWE 178  
 DB 141 TSGYLKGNAAFNVLGFRDETAADIPNVLSQAQVVELYTDTAFAWSVGARAALWE 200  
 QY 179 CGCATLGAEFOYQAQSKPKVEELNVCNVSOFSVKNPKGYKGVAFPLPTDAGVATATGKS 238  
 DB 201 CGCATLGASFQQAQSKPKVEELNVCNAAEFTINPKGYGQEFPLNINAGTVSATDTKD 260

Search completed: February 7, 2002, 21:38:10  
 Job time: 7099 sec

QY 239 ATINHEWQVGSLSYRLNSLPYIGVQWRSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 298  
 DB 261 ASIDTHEWQASLSYRLNMFTPYIGVKWSRASFDADTIRIGQPKLETSILKMTTWNPTI 320  
 QY 299 LGNATALSTTDSFSDFMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERA 358  
 DB 321 SSGSIDVDVT--KITDTLQIVSLQNLNKMKSRCGLAIGTTIVDADKYAVTVETRLIDERA 378  
 QY 359 AHVSGQFRF 367  
 DB 379 AHVNAQFRF 387







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:42:45 ; Search time 76.51 seconds  
(without alignments)  
188.811 Million cell updates/sec

Title: US-09-391-606-15

Perfect score: 2103

Sequence: 1 MLFVGNPSPDLLIDTWE.....QKLISEDLNSAVDHHHHH 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1947	92.6	389	1	OMPL_CHLPN P27455 chlamydia p
2	1882	89.5	389	1	OMIN_CHLPN Q07430 chlamydia p
3	1732	82.4	333	1	OMIK_CHLPN Q9XBf4 chlamydia p
4	1534.5	73.0	389	1	OMIA_CHLPS P16567 chlamydia p
5	1477	70.2	392	1	OMIP_CHLPS Q00087 chlamydia p
6	1415	67.3	402	1	OMLE_CHLPS P10332 chlamydia p
7	1358.5	64.6	387	1	OMPL_CHLMU P75024 chlamydia m
8	1340.5	63.7	397	1	OMIH_CHLTR P13467 chlamydia t
9	1338.5	63.6	397	1	OMIN_CHLTR P23114 chlamydia t
10	1325	63.0	396	1	OMIA_CHLTR P23732 chlamydia t
11	1315.5	62.6	393	1	OMIE_CHLTR P17451 chlamydia t
12	1315.5	62.6	397	1	OMIC_CHLTR P08780 chlamydia t
13	1313.5	62.5	393	1	OMIL_CHLTR P19542 chlamydia t
14	1308.5	62.2	393	1	OMID_CHLTR Q46409 chlamydia t
15	1307.5	62.2	395	1	OMIF_CHLTR P16155 chlamydia t
16	1295	61.6	394	1	OMIM_CHLTR P06597 chlamydia t
17	1275	60.6	394	1	OMIB_CHLTR P23421 chlamydia t
18	189	9.0	340	1	OMP2_CHLTR P38006 chlamydia t
19	182	8.7	340	1	OMP2_CHLMU Q9p113 chlamydia m
20	147.5	7.0	344	1	OMP2_CHLPN Q92752 chlamydia p
21	103	4.9	864	1	AGLU_MUCJA Q92442 mucor javan
22	100.5	4.8	919	1	HEX_ADEL2 P19900 human adeno
23	98	4.7	1848	1	CPBA_CLOCL P38058 clostridium
24	96	4.6	556	1	PHL_LEPIN P17627 leptospira
25	96	4.6	3421	1	TEGU_HSVB P28955 equine herp
26	94	4.5	448	1	FADL_ECOLI P10384 escherichia
27	93.5	4.4	2021	1	OMPA_RICCN Q52657 rickettsia
28	93.5	4.4	2292	1	POLG_EMCVB P17593 encephalomy
29	93.5	4.4	2292	1	POLG_EMCVD P17594 encephalomy
30	92.5	4.4	1169	1	YK82_YEAST P36170 saccharomyc
31	92.5	4.4	2290	1	POLG_EMCV P03304 aspergillom
32	91.5	4.4	639	1	AMYG_ASPSH P22832 aspergillus
33	90.5	4.3	1086	1	NTNM_MOUSE Q61941 mus musculu

#### RESULT 1

ID	OMPL_CHLPN	STANDARD;	PRT;	389 AA.
AC	P27455; Q9JQF6;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).			
GN	OMPA OR OMPI OR CPN0695 OR CP0051.			
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.			
OX	NCBI_TaxID=83558;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=IOL-207;			
RX	MEDLINE=91237311; PubMed=2033374;			
RA	Carter M.W., Al-Mahdawi S.A.H., Giles I.G., Trehan J.D.,			
RA	Ward M.E., Clarke I.N.;			
RT	"Nucleotide sequence and taxonomic value of the major outer membrane			
RT	protein gene of Chlamydia pneumoniae IOL-207.;"			
RL	J. Gen. Microbiol. 137:465-475(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=TWAR;			
RX	MEDLINE=91244474; PubMed=1840574;			
RA	Perez Melgosa M., Kuo C.-C., Campbell L.A.;			
RT	"Sequence analysis of the major outer membrane protein gene of			
RT	Chlamydia pneumoniae.;"			
RL	Infect. Immun. 59:2195-2199(1991).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Mitchell W.M., Tharp A.C., Stratton C.W., Sriram S.;			
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CMLO29;			
RX	MEDLINE=9206606; PubMed=10192388;			
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,			
RA	Olinger L., Grichew J., Davis R.W., Stephens R.S.;			
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.;"			
RL	Nat. Genet. 21:385-389(1999).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AR39;			
RX	MEDLINE=20150255; PubMed=10684935;			
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,			
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,			
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,			
RA	Ginn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,			
RA	Eisen J., Fraser C.M.;			
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia			
RT	pneumoniae AR39.;"			
RL	Nucleic Acids Res. 28:1397-1406(2000).			
RN	[6]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=J138;			

34	90.5	4.3	1140	1	YHJL_ECOLI	P37650 escherichia
35	90	4.3	533	1	FCY2_YEAST	P17064 saccharomyc
36	89.5	4.3	726	1	FATA_VIBAN	P11461 vibrio angu
37	89.5	4.3	867	1	SEMD_ECOLI	P77468 escherichia
38	89.5	4.3	901	1	POLG_ENMG3	P32540 mengo encep
39	89	4.2	208	1	Y396_RICPR	Q92dd5 rickettsia
40	89	4.2	363	1	YRAK_ECOLI	P43319 escherichia
41	89	4.2	895	1	Z281_HUMAN	Q9V2x9 homo sapien
42	88.5	4.2	561	1	HLVB_PROMI	P16465 proteus mir
43	88.5	4.2	639	1	AMYG_ASPAK	P23176 aspergillus
44	88.5	4.2	1035	1	MT10_YEAST	P39692 saccharomyc
45	87.5	4.2	482	1	PUR8_YEAST	Q05911 saccharomyc

#### ALIGNMENTS







QY 182 ATLGAEFOYAQSKPVEELNVICNYSQFSVKNPKGYKGVAEPLPTDAGVATATGTSATI 241  
 Db 204 ATLGAEFOYAQSKPVEELNVICNYSQFSVKNPKGYKGVAEPLPTDAGVATATGTSATI 263  
 QY 242 NYHEQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLGN 301  
 Db 264 NYHEQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLGS 323  
 QY 302 ATALSTTQSFDFMIVSQINKEKSRKACGVTGATLVADKWSLTAERLINERAHV 361  
 Db 324 ATAVSSDQSFDFMIVSQINKEKSRKACGVTGATLVADKWSLTAERLINERAHI 383  
 QY 362 SQQFRF 367  
 Db 384 SQQFRF 389

RESULT 3  
 OMIA\_CHLPSN STANDARD; PRT; 333 AA.  
 AC Q9XB4;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN (MOMP) (FRAGMENT).  
 GN OMPA OR OMP1.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiaophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KOALA TYPE I;  
 RX MEDLINE=93123168; PubMed=8419295;  
 RA Kalltenboeck B., Kousoulas K.G., Storz J.;  
 RT "Structures of and allelic diversity and relationships among the major  
 RT outer membrane protein (ompA) genes of the four chlamydial species.";  
 RL J. Bacteriol. 175:487-502(1993).  
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
 CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
 CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
 CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M73038; AAD38210.1;  
 DR InterPro; IPR000604; Chlamydia\_OMP.  
 DR Pfam; PF01308; Chlamydia\_OMP; 1.  
 DR ProDom; PD001717; Chlamydia\_OMP; 1.  
 KW Outer membrane; Transmembrane; Porin.  
 FT NON\_TER 1  
 FT NON\_TER 333  
 SQ SEQUENCE 333 AA; 35811 MW; 204604512C4C3B3F CRC64;

Query Match 82.4%; Score 1732; DB 1; Length 333;  
 Best Local Similarity 97.3%; Pred. No. 7.8e-136;  
 Matches 324; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 23 AGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKFTSMGAKTGSAAANYTTAVDRP 82  
 Db 1 AGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKFTSMGAKTGSATANYTTAVDRP 60  
 QY 83 NPAYNKHLDHAEWETNAGFIALNIWDREDFVCTLGASNGYIRGNSTAFNLVGLFGVKGT 142

Db 61 NPAYNKHLDHAEWETNAGFIALNIWDREDFVCTLGASNGYIRGNSTAFNLVGLFGVKGT 120  
 QY 143 VNANELPNVLSNGVVELYTDTSFWSVGARGALHECCGATLGAFOYAQSKPKVEELNV 202  
 Db 121 VAANELPNVLSNGVVELYTDTSFWSVGARGALHECCGATLGAFOYAQSKPKVEELNV 180  
 QY 203 ICNVQFSQFNKPKGYKGVAEPLPTDAGVATATGTSATINYHEQVQASLSYRLNSLPVY 262  
 Db 181 ICNVAQFSVKNPKGYKGVAEPLPTDAGVATATGTSATINYHEQVQASLSYRLNSLPVY 240  
 QY 263 ICVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLGNLTATSTDSFDFMIVSQI 322  
 Db 241 ICVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLGNLTATSTDSFDFMIVSQI 300  
 QY 323 NKEFKSRKACGVTGATLVADKWSLTAERLIN 355  
 Db 301 NKEFKSRKACGVTGATLVADKWSLTAERLIN 333

RESULT 4  
 OMIA\_CHLPS STANDARD; PRT; 389 AA.  
 AC P16567;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).  
 GN OMPA OR OMP1.  
 OS Chlamydia psittaci (Chlamydia psittaci).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiaophila.  
 OX NCBI\_TaxID=83554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OVINE ENZOOTIC ABORTION ISOLATE S26/3;  
 RX MEDLINE=90128177; PubMed=2612883;  
 RA Herring A.J., Tan T.W., Baxter S., Inglis N.F., Dunbar S.;  
 RT "Sequence analysis of the major outer membrane protein gene of an  
 RT ovine abortion strain of Chlamydia psittaci.";  
 RL FEMS Microbiol. Lett. 53:153-158(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BOVINE ABORTION ISOLATE BAL;  
 RX MEDLINE=96189695; PubMed=8605581;  
 RA Griffiths P.C., Plater J.M., Martin T.C., Hughes S.L.,  
 RA Hughes K.J., Hewinson R.G., Dawson M.;  
 RT "Epizootic bovine abortion in a dairy herd: characterization of a  
 RT Chlamydia psittaci isolate and antibody response.";  
 RL Br. Vet. J. 151:683-693(1995).  
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
 CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
 CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
 CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X51859; CAA36152.1;  
 DR EMBL; L39020; AAB02850.1;  
 DR PIR; S08770; MCMWP3.  
 DR InterPro; IPR000604; Chlamydia\_OMP.  
 DR Pfam; PF01308; Chlamydia\_OMP; 1.  
 DR ProDom; PD001717; Chlamydia\_OMP; 1.  
 KW Outer membrane; Transmembrane; Porin; Signal.  
 FT SIGNAL 1 22



FT CHAIN 23 389 MAJOR OUTER MEMBRANE PROTEIN.  
SQ SEQUENCE 389 AA; 41883 MW; 74185A23ACDBB447 CRC64;

Query Match 73.08; Score 1534.5; DB 1; Length 389;  
Best Local Similarity 75.7%; Pred. No. 2e-119;  
Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;

QY 2 LPVGNPSPDLLIDGTWEGAGDPCDPCATWCDAISIRAGFYGDYVDFRILKVDAPKTF 61  
DB 23 LPVGNPAEPSLLIDGTWEGASGDPDPCATWCDAISIRAGFYGDYVDFRILKVDVVKTI 82

QY 62 S-MGAKPTGSA--AANYTTAVDRNPAYNKHLDHAEWFTNAGFIALNIWDRFVCTLGASN 120  
DB 83 TGMGAVPTGTAANKYKPTDRPNIAKGLQDAEWFNAEFLALNIWDRFVCTLGASN 142

QY 121 GYIRGNSTAFNLVGLFGVKGTTVNANLNPVLSNGVVELYDTDSFWSVGARGALWECG 180  
DB 143 GYFKASSAFLNLVGLIGVKGSSIAADQLPNVIGITQGIYEFYDTDSFWSVGARGALWECG 202

QY 181 CATLGAEFQYAOQSKPKVEELNVCNVSOFVSNKPKYGVAFPLPTDAGVATATGTSAT 240  
DB 203 CATLGAEEFYAQSNPKIEMLNVSFPAQFVVKHPRGYGTAPPLPTAGTQADTTSAT 262

QY 241 INYHEWQVGLSLYRLNLSVPIYGVQSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLG 300  
DB 263 IKYHEWQVGLSLYRLNLSVPIYVNSRATFDADIRIAQPKLAAVNLTTNPTLLG 322

QY 301 NATALSTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLNERAAH 360  
DB 323 EATALDTSNKFADFLQIASIQINKMKSRKACGVAVGATLVADKWSLTAEARLNERAAH 382

QY 361 VSGQRF 367  
DB 383 MNAQRF 389

RESULT 5  
OMIE\_CHLPS  
ID OMIE\_CHLPS STANDARD; PRT; 392 AA.  
AC Q00087;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).  
GN OMPA OR OMP1.  
OS Chlamydia psittaci (Chlamydia psittaci).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. [1]  
RN NCBI\_TaxID=83554;  
RP SEQUENCE FROM N.A.  
RC STRAIN-FPN/PRING;  
RX MEDLINE=94103736; PubMed=8277245;  
RA Storey C., Lusher M., Yates P., Richmond S.;  
RT "Evidence for Chlamydia pneumoniae of non-human origin.";  
RL J. Gen. Microbiol. 139:2621-2626(1993).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X61096; CAA43409.1; -

DR PIR; A40371; A40371.  
DR PIR; S16137; S16137.  
DR InterPro: IPR000604; Chlamydia\_OMP.  
DR Pfam: PF01308; Chlamydia\_OMP; 1.  
DR ProDom: PD001717; Chlamydia\_OMP; 1.  
KW Outer membrane; Transmembrane; Porin; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 392 MAJOR OUTER MEMBRANE PROTEIN.  
SQ SEQUENCE 392 AA; 42069 MW; 88B3C5D90BBA26DB CRC64;

Query Match 70.2%; Score 1477; DB 1; Length 392;  
Best Local Similarity 74.1%; Pred. No. 1.1e-114;  
Matches 275; Conservative 38; Mismatches 52; Indels 6; Gaps 4;

QY 2 LPVGNPSPDLLIDGTWEGAGDPCDPCATWCDAISIRAGFYGDYVDFRILKVDAPKTF 61  
DB 23 LPVGNPAEPSLLIDGTWEGASGDPDPCATWCDAISIRAGFYGDYVDFRILKVDVVKTI 82

QY 62 S-MGAKPTGSA--AANYTTAVDRNPAYNKHLDHAEWFTNAGFIALNIWDRFVCTLGA 118  
DB 83 SGMAAAPTAASTASNTTVAADRSNFAYKHLQDAEWCNTNAAYLALNIWDRFVCTLGA 142

QY 119 SNGYIRGNSTAFNLVGLFGVKGTTVNANLNPVLSNGVVELYDTDSFWSVGARGALWE 178  
DB 143 SNGYFKASSDAFLNLVGLIGLAGTDF--ANQRPNVETISQIVELYDTDFAFWSVGARGALWE 201

QY 179 CGCATLGAEFQYAOQSKPKVEELNVCNVSOFVSNKPKYGVAF--FLPTDAGVATATGT 236  
DB 202 CGCATLGAEFQYAOQSNPKIEMLNVTSSPAQFMIHPRGYGTAAFPPLPVAAAGTATATDT 261

QY 237 KSAITINHEWQVGLSLYRLNLSVPIYGVQSRATFDADNIRIAQPKLPTAVLNLTAWN 296  
DB 262 KSAITVHEWQVGLSLYRLNLSVPIYGVNWSRATFDADTIRIAQPKLASAILNLTWN 321

QY 297 SLLGNATALSTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLINE 356  
DB 322 TLLGVATTLDTSNKFADFLQIASIQINKMKSRKACGVAVGATLVADKWSLTAEARLIDE 381

QY 357 RAAHVSGQRF 367  
DB 382 RAAHINAQRF 392

RESULT 6  
OMIE\_CHLPS  
ID OMIE\_CHLPS STANDARD; PRT; 402 AA.  
AC P10332;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).  
GN OMPA OR OMP1.  
OS Chlamydia psittaci (Chlamydia psittaci).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. [1]  
RN NCBI\_TaxID=83554;  
RP SEQUENCE FROM N.A.  
RC STRAIN-EAE A22/W;  
RA Pickett M.A., Everson S.J., Clarke I.N.;  
RT "Chlamydia psittaci ewe abortion agent: complete nucleotide sequence of the major outer membrane protein gene.";  
RL FEMS Microbiol. Lett. 55:229-234(1988).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -







```
SQ SEQUENCE 387 AA; 42009 MW; 4FD6DC23248E0A2 CRC64;
Query Match 64.6%; Score 1358.5; DB 1; Length 387;
Best Local Similarity 66.7%; Pred. No. 6.8e-105;
Matches 246; Conservative 50; Mismatches 66; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLIDGTIWEGAGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSPMLIDGILWEGFGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDVKNQF 82
QY 62 SMGAKPTGSAANYTTA---VDRENPAYNKHLHDAEFTNAGFIALNLTWDFRDFVCTLG 118
DB 83 EMGAAPTGD--ADLTAPTPASRENPAYGKHMQDAEMFTNAAYNMALNIWDFRDFVCTLGA 140
QY 119 SNGYIRGNSTAFNLVGLFGVGTNNANLPSVLSNGVVELYDTSFSSWVGARGALWE 178
DB 141 TSGYLKGNAAFNLVGLFGVGTNNANLPSVLSNGVVELYDTSFSSWVGARGALWE 200
QY 179 CGCATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYGVKGVAPPLPTDAGVATATGTS 238
DB 201 CGCATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYGVKGVAPPLPTDAGVATATGTS 260
QY 239 ATINYHEQVQASLYRLNSLVPYIGVOWSRATEDADNIRIAQKLPKPTAVLNLTAWNPSL 298
DB 261 ASIDYHEQVQASLYRLNSLVPYIGVOWSRATEDADNIRIAQKLPKPTAVLNLTAWNPSL 320
QY 299 LGNATALSTTDSFDMOIVSCQINKEFKSRACGVTVGATLVADKWSLTAEARLINERA 358
DB 321 SGSGIDVDT--KIDTTLQIVSLQNLKMSKSCGLAIGTTIIVADKYAVTVETRLIDERA 378
QY 359 AHVSGQFRF 367
DB 379 AHVNAQFRF 387

RESULT 8
OMIH_CHLTR STANDARD; PRT; 397 AA.
AC P13467;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR H PRECURSOR (MOMP).
GN OMPA OR OMP13.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROVAR H;
RX MEDLINE=90045958; PubMed=2813066;
RA Hamilton P.T., Malinowski D.P.;
RT "Nucleotide sequence of the major outer membrane protein gene from Chlamydia trachomatis serovar H.";
RL Nucleic Acids Res. 17:8366-8366(1989).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=H/UW-4;
RX Dean D.A.;
RT "Sequence analysis of the major outer membrane protein gene (ompA) of Chlamydia trachomatis.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16007; CA34145.1; -.
DR EMBL; AF304857; AAG41415.1; -.
DR PIR; S06589; MNCWTH; Chlamydia_OMP.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 22
FT CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR H.
SQ SEQUENCE 397 AA; 42946 MW; 478ACE380BF37BA CRC64;
Query Match 63.7%; Score 1340.5; DB 1; Length 397;
Best Local Similarity 65.1%; Pred. No. 2.2e-103;
Matches 244; Conservative 51; Mismatches 71; Indels 9; Gaps 3;

QY 2 LPVGNPSPSLIDGTIWEGAGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSPMLIDGILWEGFGDPCDPCATWCDALISMRVGYGDFVDFRVLKTDVKNKEF 82
QY 62 SMGAKPTGSAANY---TTAVDRPNPAYNKHLHDAEFTNAGFIALNIWDFRDFVCTLG 117
DB 83 OMGAAPTNDAAIDQNDPKTNVARNPAYGKHMQDAEMFTNAAYNMALNIWDFRDFVCTLG 142
QY 118 ASNGYIRGNSTAFNLVGLFGVGTNNANL---LPNYSLSNGVVELYDTSFSSWVGARG 174
DB 143 ATTGYLKGNASFNVLVGLFGVGTNNANL---LPNYSLSNGVVELYDTSFSSWVGARG 202
QY 175 ALWECGCATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYGVKGVAPPLPTDAGVATAT 234
DB 203 ALWECGCATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYGVKGVAPPLPTDAGVATAT 262
QY 235 GTSKATINYHEQVQASLYRLNSLVPYIGVOWSRATEDADNIRIAQKLPKPTAVLNLTAW 294
DB 263 GTSKATINYHEQVQASLYRLNSLVPYIGVOWSRATEDADNIRIAQKLPKPTAVLNLTAW 322
QY 295 NPSLLGNAT--ALSTTDSFDMOIVSCQINKEFKSRACGVTVGATLVADKWSLTAEAR 352
DB 323 NPTIAGKGVTVVSGSDNDLADTMQIVSLQNLKMSKSCGLAIGTTIIVADKYAVTVETR 382
QY 353 LINERAAHVSGQFRF 367
DB 383 LIDERRAAHVNAQFRF 397

RESULT 9
OMIH_CHLTR STANDARD; PRT; 397 AA.
AC P23114;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3 PRECURSOR (MOMP).
GN OMPA OR OMP13.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=404 / SEROVAR L3;
RX MEDLINE=91285439; PubMed=2060793;
RA Fielder T.J., Peterson E.M., de la Maza L.M.;
RT "Nucleotide sequence of DNA encoding the major outer membrane protein of Chlamydia trachomatis serovar L3.";
RL Gene 101:159-160(1991).
CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
```



CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
 CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; X55700; CA39226.1; -  
 CC PIR; JE0413; JE0413.  
 CC InterPro; IPR000604; Chlamydia\_OMP.  
 CC Pfam; PF01308; Chlamydia\_OMP; 1.  
 CC ProDom; PD001717; Chlamydia\_OMP; 1.  
 CC Outer membrane; Transmembrane; Porin; Signal.  
 CC SIGNAL 1 22 BY SIMILARITY.  
 CC CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3.  
 CC SEQUENCE 397 AA; 42885 MW; FIDDCF09535C2595 CRC64;

DR EMBL; X55700; CA39226.1; -  
 DR PIR; JE0413; JE0413.  
 DR InterPro; IPR000604; Chlamydia\_OMP.  
 DR Pfam; PF01308; Chlamydia\_OMP; 1.  
 DR ProDom; PD001717; Chlamydia\_OMP; 1.  
 DR Outer membrane; Transmembrane; Porin; Signal.  
 DR SIGNAL 1 22 BY SIMILARITY.  
 DR CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3.  
 DR SEQUENCE 397 AA; 42885 MW; FIDDCF09535C2595 CRC64;

Query Match 63.6%; Score 1338.5; DB 1; Length 397;  
 Best Local Similarity 65.1%; Pred. No. 3.1e-103;  
 Matches 244; Conservative 49; Mismatches 73; Indels 9; Gaps 3;

QY 2 LPVGNPSPSLLIDGTIWEAGAAGDPCDPCATWCDALISLRAGFYGDYFDRILKVDAPKTF 61  
 DB 23 LPVGNPAEPSPMLIDGILWEGFGDPCDPCATWCDALISLRAGFYGDYFDRILKVDAPKTF 82  
 QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKLHDAEFTNAGFTALNIWDRFVFCPLG 117  
 DB 83 QMGAEPTSDTAGLNDPTTNAVARNPAYGKHMDAEMFTNAAYNALNIWDRFVFCPLG 142  
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGTTVANE---LPNVSLSNGVELYTDTSFWSVGARG 174  
 DB 143 ATTGYLKGNSAFNLVGLFGVKGTTSTNFNTAKLPNTALNQAVVELYTDTSFWSVGARG 202  
 QY 175 ALWECGCATLGAEOYQAKPKVBEELNVICNVSOFSVKNPKGYKGVAPPLPTDAGVATAT 234  
 DB 203 ALWECGCATLGAEOYQAKPKVBEELNVICNVSOFSVKNPKGYKGVAPPLPTDAGVATAT 262  
 QY 235 GTSKATINHEWQASLSYRLNSLVPIYGVQWSRATFDADNIRIAQPKLPTAVLNLTAW 294  
 DB 263 GTKDASIDYHEWQASLSYRLNSLVPIYGVQWSRATFDADNIRIAQPKLPTAVLNLTAW 322  
 QY 295 NPSLL--GNATALSTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAPAR 352  
 DB 323 NPTAGKGVSVASGSENELADTMQIVSLQNLNKMRSKSGIAVGTTIVDADKYAVTVETR 382  
 QY 353 LINERAAHVSGQFRF 367  
 DB 383 LIDERAAHVNAQFRF 397

RESULT 10  
 OM1A\_CHLTR STANDARD; PRT; 396 AA.  
 ID OM1A\_CHLTR  
 AC P23732;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR A PRECURSOR (MOMP).  
 GN OMPA OR OMPIA.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sal/OT / SEROVAR A;

RX MEDLINE=91045088; PubMed=2235504;  
 RA Hayes L.J., Clarke I.N.;  
 RT "Nucleotide sequence of the major outer membrane protein gene of  
 Chlamydia trachomatis strain A/Sal/OT";  
 RL Nucleic Acids Res. 18:6136-6136(1990).  
 CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
 CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
 CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
 CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; M58938; AAA23141.1; -  
 CC EMBL; M33635; AAA92785.1; -  
 CC PIR; SI2799; SI2799.  
 CC InterPro; IPR000604; Chlamydia\_OMP.  
 CC Pfam; PF01308; Chlamydia\_OMP; 1.  
 CC ProDom; PD001717; Chlamydia\_OMP; 1.  
 CC Outer membrane; Transmembrane; Porin; Signal.  
 CC SIGNAL 1 22  
 CC CHAIN 23 396 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR A.  
 CC SEQUENCE 396 AA; 42877 MW; 2F9D3B0CE2D08162 CRC64;

Query Match 63.0%; Score 1325; DB 1; Length 396;  
 Best Local Similarity 64.7%; Pred. No. 4.1e-102;  
 Matches 242; Conservative 47; Mismatches 77; Indels 8; Gaps 3;

QY 2 LPVGNPSPSLLIDGTIWEAGAAGDPCDPCATWCDALISLRAGFYGDYFDRILKVDAPKTF 61  
 DB 23 LPVGNPAEPSPMLIDGILWEGFGDPCDPCATWCDALISLRAGFYGDYFDRILKVDAPKTF 82  
 QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKLHDAEFTNAGFTALNIWDRFVFCPLG 117  
 DB 83 QMGAEPTSDTAGLNDPTTNAVARNPAYGKHMDAEMFTNAAYNALNIWDRFVFCPLG 142  
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGTTV---NANELPNVSLNSGVVELYTDTSFWSVGARG 174  
 DB 143 ATTGYLKGNSAFNLVGLFGVKGTTSTNFNTAKLPNTALNQAVVELYTDTSFWSVGARG 202  
 QY 175 ALWECGCATLGAEOYQAKPKVBEELNVICNVSOFSVKNPKGYKGVAPPLPTDAGVATAT 234  
 DB 203 ALWECGCATLGAEOYQAKPKVBEELNVICNVSOFSVKNPKGYKGVAPPLPTDAGVATAT 262  
 QY 235 GTSKATINHEWQASLSYRLNSLVPIYGVQWSRATFDADNIRIAQPKLPTAVLNLTAW 294  
 DB 263 GTKDASIDYHEWQASLSYRLNSLVPIYGVQWSRATFDADNIRIAQPKLPTAVLNLTAW 322  
 QY 295 NPSLLGNATALSTDS--FSDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARL 353  
 DB 323 NPTAGKGVSVASGSENELADTMQIVSLQNLNKMRSKSGIAVGTTIVDADKYAVTVETR 382  
 QY 354 INERAAHVSGQFRF 367  
 DB 383 IDERAAHVNAQFRF 396

RESULT 11  
 OM1E\_CHLTR STANDARD; PRT; 393 AA.  
 ID OM1E\_CHLTR  
 AC P17451;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)







Db 203 ALWECGCATLGASFOYAQSKPKVEELNVLNASEFTINKPKYGVGAEPFLNITAGTEAAT 262  
Qy 235 GTSKATINHEWQVGSASLYRLNSLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 294  
Db 263 GTSKASIDYHEWQASLALSRLNFTPIGVKWSRVSFADATIRIAQPKLAEILDVTTL 322  
Qy 295 NPSLLGNATALS--TTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEAR 352  
Db 323 NRTTAGKSVVSAGTDELADMTQIVSLQNLNKKSRKSCGIAVGTTIVDAKYAVTVEAR 382  
Qy 353 LINERAAHVSGQFRF 367  
Db 383 LIDERRAAHVNAQFRF 397

## RESULT 13

OM1L\_CHLTR  
ID OM1L\_CHLTR STANDARD; PRT; 393 AA.  
AC P19542;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L1 PRECURSOR (MOMP).  
GN OMPA OR OMP1L.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Pickett M.A., Ward M.E., Clarke I.N.;  
RT "Complete nucleotide sequence of the major outer membrane protein  
gene from Chlamydia trachomatis serovar L1.";  
RL FEMS Microbiol. Lett. 42:185-190(1987).  
CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M36533; AAA23142.1; -.  
DR PIR; S06259; S06259.  
DR InterPro; IPR000604; Chlamydia\_OMP.  
DR Pfam; PF01308; Chlamydia\_OMP; 1.  
DR ProDom; PD00117; Chlamydia\_OMP; 1.  
KW Outer membrane; Transmembrane; Porin; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L1.  
SQ SEQUENCE 393 AA; 42543 MW; 7A952839408EE2DF CRC64;

Query Match 62.58; Score 1313.5; DB 1; Length 393;  
Best Local Similarity 64.58; Pred. No. 3.6e-101;  
Matches 240; Conserved 51; Mismatches 74; Indels 7; Gaps 4;

Qy 2 LPVGNPSPSLIDGTIWEAGAGPCDPCATWCDAISLRAGFYGVYDFDRILKYDPAKTF 61  
Db 23 LPVGNPAEPLMIDGILLWEGFGDPCCTTWCDAISRMGYGVDFVDRVLQTDVNEKF 82  
Qy 62 SMGAKP----TGSAAANYTAVDRPNPAYNKLHDAEWFTNAGFTALNIWDRFVFCFLGA 118  
Db 83 QMGAKPTATTGNAAP-STCTARENPAYGRHMQDAEMFTNAAYNALNIWDRFVFCFLGA 141  
Qy 119 SNGYIRGNSTAFNLVGLFG--VKGTTVNANLNPVLSLNGVVELYTDTSFSWSVGARGAL 176

Db 142 TSGLYKGNASFNVLGFLGDNENQSTVKKDAVPNMSFDQSVVELYTDITTFMSVGARAAL 201  
Qy 177 WECCATILGAEFOYAQSKPKVEELNVLNASEFTINKPKYGVGAEPFLNITAGTEAAT 236  
Db 202 WECCATILGASFOYAQSKPKVEELNVLNASEFTINKPKYGVGAEPFLNITAGTEAAT 261  
Qy 237 KEATINHEWQVGSASLYRLNSLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 296  
Db 262 KASIDYHEWQASLALSRLNFTPIGVKWSRVSFADATIRIAQPKLAEILDVTTL 321  
Qy 297 NPSLLGNATALS--TTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEAR 355  
Db 323 NRTTAGKSVVSAGTDELADMTQIVSLQNLNKKSRKSCGIAVGTTIVDAKYAVTVEAR 381  
Qy 353 LINERAAHVSGQFRF 367  
Db 383 LIDERRAAHVNAQFRF 397

## RESULT 14

OM1D\_CHLTR  
ID OM1D\_CHLTR STANDARD; PRT; 393 AA.  
AC Q46409;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR D PRECURSOR (MOMP).  
GN OMPA OR OMP1 OR CT681.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=D/B-120;  
RC MEDLINE=93013014; PubMed=1398119;  
RA Sayada C., Denamur E., Elion J.;  
RT "Complete sequence of the major outer membrane protein-encoding gene  
of Chlamydia trachomatis serovar Da.";  
RL Gene 120:129-130(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA STRAIN=D/1U-71960;  
RC MEDLINE=98339860; PubMed=9673241;  
RA Stothard D.R., Boguslawski G., Jones R.B.;  
RT "Phylogenetic analysis of the Chlamydia trachomatis major outer  
membrane protein and examination of potential pathogenic  
determinants.";  
RL Infect. Immun. 66:3618-3625(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA STRAIN=D/UW-3/CX;  
RC MEDLINE=99000809; PubMed=9784136;  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis.";  
RL Science 282:754-759(1998).  
CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----



or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL; X62918; CAA44701.1; -  
 CC EMBL; AF063195; AAC31436.2; -  
 CC EMBL; AE001338; AAC68276.1; -  
 CC InterPro; IPR000604; Chlamydia\_OMP.  
 CC Pfam; PF01308; Chlamydia\_OMP; 1.  
 CC ProDom; PD001717; Chlamydia\_OMP; 1.  
 CC Outer membrane; Transmembrane; Porin; Signal; Complete proteome.  
 KW SIGNAL 1 22 BY SIMILARITY.  
 FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR D.  
 SQ SEQUENCE 393 AA; 42438 MW; 8CD692FD3EEF21D6 CRC64;

Query Match 62.2%; Score 1308.5; DB 1; Length 393;  
 Best Local Similarity 64.8%; Pred. No. 9.3e-101;  
 Matches 241; Conservative 48; Mismatches 76; Indels 7; Gaps 4;  
 QY 2 LPVGNPDPSSLIDGTIWEAGAAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 DB 23 LPVGNPAEPSLMDIGILWEGFGDPCPCATWCDAISMRVGYGDFVDRVLKTDVKNKEF 82  
 QY 62 SMGAKP---TGSAANTTAVDRPNPAYNKHLDHDAEFTNAGFTALNIWDRFDFVCTLGA 118  
 DB 83 QMGAKPTTDTGNSAAP-STLTARENPAYGRHMQDAEFTNAACMALNIWDRFDFVCTLGA 141  
 QY 119 SNGYIRGNSTAFNLVGLFG--VKGTTYNANLNPVNSLNGVVELYTDTSFWSVGARGAL 176  
 DB 142 TSGYLKNSASFNLVGLFGDGNENQKTVKAEVPSNNSFQSVVELYTDTSFWSVGARGAL 201  
 QY 177 WECGCATLGAEFOYAQSKPKVEELNVCNVSQFSVNRKPKYKGVAFFLPTDAGVATATGT 236  
 DB 202 WECGCATLGAEFOYAQSKPKVEELNVCNVAEFTINKPKYGVGKEFFLPTDAGVATATGT 261  
 QY 237 KSAITNHEWQVGSASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWN 296  
 DB 262 KASIDYHEWQASLSYRLNMFPTPIGVKWSRASFDADTIRIAQPKSATAIFDTTLNP 321  
 QY 297 SLGNATALSTTD--SFSDFMQIVSCQINKFKSKACGVTVGATLVADKWSLTAEARLI 355  
 DB 322 TIAGAGVKYGAEGQLGDTMQIVSLQNLNKKSRKSCGIAGVTIVDADKAVYVETRLID 381  
 QY 356 ERAAHVSGQFRF 367  
 DB 382 ERAAHVNAQFRF 393

RESULT 15  
 OMIF\_CHLTR STANDARD; PRT; 395 AA.  
 AC P16155;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR F PRECURSOR (MOMP).  
 GN OMPA OR OMPF.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IC-CAL3 / SEROVAR F;  
 RX MEDLINE=90192102; PubMed=2315025;  
 RA Zhang Y.X., Morrison S.G., Caldwell H.D.;  
 RT "The nucleotide sequence of major outer membrane protein gene of  
 Chlamydia trachomatis serovar F.";  
 RL Nucleic Acids Res. 18:1061-1061(1990).  
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
 BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
 THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
 MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.

CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X52080; CAA36299.1; -  
 CC PIR; S08463; MNCWTF.  
 CC InterPro; IPR000604; Chlamydia\_OMP.  
 CC Pfam; PF01308; Chlamydia\_OMP; 1.  
 CC ProDom; PD001717; Chlamydia\_OMP; 1.  
 KW Outer membrane; Transmembrane; Porin; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 395 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR F.  
 SQ SEQUENCE 395 AA; 42586 MW; 7F90FFDEEC264ACF CRC64;

Query Match 62.2%; Score 1307.5; DB 1; Length 395;  
 Best Local Similarity 63.5%; Pred. No. 1.1e-100;  
 Matches 237; Conservative 53; Mismatches 76; Indels 7; Gaps 3;  
 QY 2 LPVGNPDPSSLIDGTIWEAGAAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 DB 23 LPVGNPAEPSLMDIGILWEGFGDPCPCITWCDAISMRMGYGGDFVDRVLKTDVKNKEF 82  
 QY 62 SMG---AKPTGSAANTTAVDRPNPAYNKHLDHDAEFTNAGFTALNIWDRFDFVCTLGA 118  
 DB 83 EMGEALAGASGNTTSLTKLVERTNPAYGKHMQDAEFTNAACMTLNIWDRFDFVCTLGA 142  
 QY 119 SNGYIRGNSTAFNLVGLF--GVKGTTYNANLNPVNSLNGVVELYTDTSFWSVGARGAL 176  
 DB 143 TSGYLKNSASFNLVGLFGDGNENQKTVKAEVPSNNSFQSVVELYTDTSFWSVGARGAL 202  
 QY 177 WECGCATLGAEFOYAQSKPKVEELNVCNVSQFSVNRKPKYKGVAFFLPTDAGVATATGT 236  
 DB 203 WECGCATLGAEFOYAQSKPKVEELNVCNVAEFTINKPKYGVGKEFFLPTDAGVATATGT 262  
 QY 237 KSAITNHEWQVGSASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWN 296  
 DB 263 KASIDYHEWQASLSYRLNMFPTPIGVKWSRASFDSDTIRIAQPKRLVTPVVDITLNP 322  
 QY 297 SL--LGNATALSTTDSFSDFMQIVSCQINKFKSKACGVTVGATLVADKWSLTAEARLI 354  
 DB 323 TIAGCGSVAGANTEGQISDTMQIVSLQNLNKKSRKSCGIAGVTIVDADKAVYVETRLI 382  
 QY 355 NERAAHVSGQFRF 367  
 DB 383 DERAHVNAQFRF 395

Search completed: February 7, 2002, 21:42:47  
 Job time: 601 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:41:19 ; Search time 172 Seconds  
(without alignments)  
335.066 Million cell updates/sec

Title: US-09-391-606-15  
Perfect score: 2103  
Sequence: 1 MLPVGNPSPDLLIDTWE.....OKLISEEDLNSAVDHHHHH 394

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_17.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1910	90.8	389	Q08085	Q08085 chlamydia p
2	1531.5	72.8	389	Q9APM4	Q9APM4 chlamydophi
3	1529.5	72.7	388	Q9AIK1	Q9AIK1 chlamydia p
4	1501.5	71.4	391	Q46235	Q46235 chlamydia p
5	1497.5	71.2	388	Q9AIK0	Q9AIK0 chlamydia p
6	1497.5	71.2	389	Q9AIH9	Q9AIH9 chlamydophi
7	1487	70.7	392	Q9AIJ4	Q9AIJ4 chlamydia p
8	1480	70.4	381	Q9AIJ2	Q9AIJ2 chlamydophi
9	1479	70.3	390	Q9AIJ5	Q9AIJ5 chlamydia p
10	1477	70.2	392	Q99Q80	Q99Q80 chlamydophi
11	1469.5	69.9	380	Q9AIJ1	Q9AIJ1 chlamydophi
12	1423	67.7	391	Q9AIJ2	Q9AIJ2 chlamydia p
13	1420.5	67.5	341	Q9X717	Q9X717 chlamydophi
14	1419	67.5	402	Q46193	Q46193 chlamydia p
15	1415	67.3	382	Q9AIJ9	Q9AIJ9 chlamydia p
16	1415	67.3	395	Q9AIJ7	Q9AIJ7 chlamydia p
17	1415	67.3	397	Q9AIJ8	Q9AIJ8 chlamydia p
18	1413	67.2	402	Q9AIJ6	Q9AIJ6 chlamydia p
19	1410	67.0	402	Q9AIJ0	Q9AIJ0 chlamydia p

20	1408	67.0	402	2	Q46203	Q46203 chlamydia p
21	1407	66.9	402	2	Q46236	Q46236 chlamydia p
22	1382	65.7	402	2	Q9AIJ3	Q9AIJ3 chlamydia p
23	1371.5	65.2	352	2	Q70085	Q70085 chlamydia p
24	1367.5	65.0	352	2	Q70050	Q70050 chlamydia p
25	1363.5	64.8	352	2	Q69307	Q69307 chlamydia p
26	1362.5	64.8	352	2	Q69306	Q69306 chlamydia p
27	1357	64.5	337	2	Q9XBP6	Q9XBP6 chlamydophi
28	1353	64.3	336	2	Q9XBP5	Q9XBP5 chlamydophi
29	1353	64.3	353	2	Q69305	Q69305 chlamydia p
30	1352	64.3	396	2	Q9AIJ8	Q9AIJ8 chlamydia s
31	1351	64.2	404	2	Q46407	Q46407 chlamydia t
32	1343	63.9	389	2	Q9AIJ4	Q9AIJ4 chlamydia s
33	1342	63.8	396	2	Q69093	Q69093 chlamydia t
34	1341.5	63.8	386	2	Q9AIJ5	Q9AIJ5 chlamydia s
35	1339.5	63.7	397	2	Q69094	Q69094 chlamydia t
36	1337.5	63.6	397	2	Q69095	Q69095 chlamydia t
37	1336.5	63.6	387	2	Q9AIJ0	Q9AIJ0 chlamydia s
38	1335.5	63.5	397	2	Q9F951	Q9F951 chlamydia t
39	1333	63.4	340	2	Q9XBP2	Q9XBP2 chlamydophi
40	1331.5	63.3	387	2	Q9AIJ1	Q9AIJ1 chlamydia s
41	1331.5	63.3	397	2	Q9F950	Q9F950 chlamydia t
42	1330.5	63.3	385	2	Q9AIJ6	Q9AIJ6 chlamydia s
43	1328	63.1	396	2	Q46406	Q46406 chlamydia t
44	1327.5	63.1	385	2	Q9AIJ7	Q9AIJ7 chlamydia s
45	1326	63.1	356	2	Q52924	Q52924 chlamydia p

## ALIGNMENTS

RESULT 1

Q08085 PRELIMINARY; PRT; 389 AA.  
AC Q08085;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).  
DE "Remarkable sequence relatedness in the DNA encoding the major outer  
OS Chlamydia psittaci (Chlamydophila psittaci).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
OX NCBI\_TaxID=83554;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KOALA TYPE 1;  
RX MEDLINE=94171025; PubMed=8125292;  
RA Girjes A.A., Carrick F.N., Lavin M.F.;  
RT "Remarkable sequence relatedness in the DNA encoding the major outer  
RT membrane protein of Chlamydia psittaci (koala type 1) and Chlamydia  
RT pneumoniae.";  
RL Gene 138:139-142(1994).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
DR EMBL; J2023; CAA50906.1;  
DR InterPro; IPR000604; Chlamydia\_OMP.  
DR Pfam; PR01308; Chlamydia\_OMP; 1.  
DR PRINTS; PR01334; CHLAMIDIADOMP.  
DR ProDom; PD001717; Chlamydia\_OMP; 1.  
DR Outer membrane; Transmembrane; Porin; Signal.  
FT SIGNAL 1 23 BY SIMILARITY.  
FT CHAIN 24 389 MAJOR OUTER MEMBRANE PROTEIN.  
SQ SEQUENCE 389 AA; 41579 MW; 5DC50E85A6F4E50F CRC64;

Query Match 90.8%; Score 1910; DB 2; Length 389;  
Best Local Similarity 97.5%; Pred. No. 2,1e-146;  
Matches 357; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 LPVGNPSPDLLIDTWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRLKVDAPKTF 61



```

Db 24 LPVGNPDPSSLIDGTWEGAAGDPCDPCATWCDAISLRAGFYGYVDFRILKVDAPKTF 83
QY 62 SMGAKPTGSAANNTTAVDRPNPYNKHLHDAEFTNAGFIALNIWDRFDVFTCLGASNG 121
Db 84 SMGAKPTGSAANTTAVDRPNPYNKHLHDAEFTNAGFIALNIWDRFDVFTCLGASNG 143
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANLPNVSLNSGVVVELYDTDSFSVSGARGALWECG 181
Db 144 YIRGNSTAFNLVGLFGVKGTTVNANLPNVSLNSGVVVELYDTDSFSVSGARGALWECG 203
QY 182 ATLGAEFQYAGQSPKVEELNVICNVQSVNPKPGYKGVAFPLPTDAGVATATGKTSATI 241
Db 204 ATLGAEFQYAGQSPKVEELNVICNVQSVNPKPGYKGVAFPLPTDAGVATATGKTSATI 263
QY 242 NYHEWQVGLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLGN 301
Db 264 NYHEWQVGLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLGN 323
QY 302 ATALSTTDSFDFMOIYVSCQINKFKSRKACGTVTGATLVDAWKSLTAEARLINERAHV 361
Db 324 TTTATSDSFDFMOIYVSCQINKFKSRKACGTVTGATLVDAWKSLTAEARLINERAHV 383
QY 362 SGQFRF 367
Db 384 SGQFRF 389

RESULT 2
Q9APM4
ID Q9APM4 PRELIMINARY; PRT; 389 AA.
AC Q9APM4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMP1.
OS Chlamydia abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LLG;
RX MEDLINE=20569239; PubMed=11119563;
RA Vreton E., Psarrou E., Kalsar M., Vlisidou I., Salti-Montesanto V.,
  Longbottom D.;
RT "Identification of protective epitopes by sequencing of the major
  outer membrane protein gene of a variant strain of Chlamydia psittaci
  serotype 1."
RL Infect. Immun. 69:607-612(2001).
DR EMBL; AF272945; AAG53881.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN 1 22 POTENTIAL.
FT CHAIN 23 389 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 389 AA; 41897 MW; 20513C69C7DBAAF5 CRC64;

Query Match 72.8%; Score 1531.5; DB 2; Length 389;
Best Local Similarity 75.7%; Pred. No. 8.2e-116;
Matches 278; Conservative 38; Mismatches 50; Indels 1; Gaps 1;

QY 2 LPVGNPDPSSLIDGTWEGAAGDPCDPCATWCDAISLRAGFYGYVDFRILKVDAPKTF 61
Db 23 LPVGNPAEPLSLIDGTWEGASGDPDPCSTWCDAISIRAGYGYGVDFRVLKVDVNTKI 82
QY 62 S-MGAKPTGSAANNTTAVDRPNPYNKHLHDAEFTNAGFIALNIWDRFDVFTCLGASN 120
Db 83 TMGAVPTGTAAADYKTPTRNPNTAYGKHLQDAEFTNAAFLNALNIWDRFDVFTCLGASN 142
QY 121 YIRGNSTAFNLVGLFGVKGTTVNANLPNVSLNSGVVVELYDTDSFSVSGARGALWECG 180
Db 143 GYFKASSAFLNLVGLIGVKGTSVAADQLPNVGITGIVEFTDTTFSVSGARGALWECG 202

```

```

QY 181 CATLGAEFQYAGQSPKVEELNVICNVQSVNPKPGYKGVAFPLPTDAGVATATGKTSAT 240
Db 203 CATLGAEFQYAGQSPKVEELNVICNVQSVNPKPGYKGVAFPLPTDAGVATATGKTSAT 262
QY 241 INYHEWQVGLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLG 300
Db 263 IKYHEWQVGLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLG 322
QY 301 NATALSTTDSFDFMOIYVSCQINKFKSRKACGTVTGATLVDAWKSLTAEARLINERAHV 360
Db 323 EATLTLDNFKFAFPLQIASIQINKMKSRKACGVAAGATLIDADKWSITGEARLINERAHV 382
QY 361 VSGQFRF 367
Db 383 MNAQFRF 389

RESULT 3
Q9AIK1
ID Q9AIK1 PRELIMINARY; PRT; 388 AA.
AC Q9AIK1;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VS225;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae."
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL; AF269259; AAK0240.1; -.
KW Signal.
FT NON_TER 1 1 POTENTIAL.
FT SIGNAL <1 19
FT CHAIN 20 388 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 388 AA; 41573 MW; 8E23D22C9B9948D CRC64;

Query Match 72.7%; Score 1529.5; DB 2; Length 388;
Best Local Similarity 76.2%; Pred. No. 1.2e-115;
Matches 281; Conservative 35; Mismatches 50; Indels 3; Gaps 2;

QY 2 LPVGNPDPSSLIDGTWEGAAGDPCDPCATWCDAISLRAGFYGYVDFRILKVDAPKTF 61
Db 20 LPVGNPAEPLSLIDGTWEGASGDPDPCATWCDAISIRAGYGYGVDFRVLKVDVNTKI 79
QY 62 S-MGAKPTGSAANNTTAVDRPNPYNKHLHDAEFTNAGFIALNIWDRFDVFTCLGASN 120
Db 80 SGMAAPTGSAAADYKTPTRNPNTAYGKHLQDAEFTNAAFLNALNIWDRFDVFTCLGASN 139
QY 121 YIRGNSTAFNLVGLFGVKGTTVNANLPNVSLNSGVVVELYDTDSFSVSGARGALWECG 180
Db 140 GYFKASSAFLNLVGLIGVKGTSVAADQLPNVGITGIVEFTDTTFSVSGARGALWECG 199
QY 181 CATLGAEFQYAGQSPKVEELNVICNVQSVNPKPGYKGVAFPLPTDAGVATATGKTS 238
Db 200 CATLGAEFQYAGQSPKVEELNVICNVQSVNPKPGYKGVAFPLPTDAGVATATGKTS 259
QY 239 ATINYHEWQVGLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 298
Db 260 ATLKYHEWQVGLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 319
QY 299 LGNATALSTTDSFDFMOIYVSCQINKFKSRKACGTVTGATLVDAWKSLTAEARLINERA 358
Db 320 LGNATALSTTDSFDFMOIYVSCQINKFKSRKACGTVTGATLVDAWKSLTAEARLINERA 379
QY 359 AHVSGQFRF 367

```







RESULT	7	
Q9AIJ4		
ID	Q9AIJ4	PRELIMINARY; PRT; 392 AA.
AC	Q9AIJ4;	
CD	01-JUN-2001 (TREMBlrel. 17, Created)	
DD	01-JUN-2001 (TREMBlrel. 17, Last sequence update)	
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)	
DE	01-JUN-2001 (TREMBlrel. 17, Last annotation update)	
DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.	
OMPA.		
OS	Chlamydia psittaci (Chlamydophila psittaci).	
OS	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.	
OC	NCBI_TaxID=83554;	
OC	[1]	
RRN	SEQUENCE FROM N.A.	
RRP	STRAIN=TEXAS TURKEY 3, TT3.	
RRC	MEDLINE=21078680; PubMed=11211261;	
RRA	Bush R.M., Everett K.D.;	
RXX	"Molecular evolution of the Chlamydiaceae."	
RL	Int. J. Syst. Evol. Microbiol. 51:203-220(2001).	

RA Everett K.D.E., Hamblly W.A., Andersen A.A.;  
 RRL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF269279; AAK00260.1; -.



```

KW Signal. 1 15 POTENTIAL.
FT NON_TER <1 15
FT SIGNAL 16 381 MAJOR OUTER MEMBRANE PROTEIN.
FT CHAIN 16 381
SQ SEQUENCE 381 AA; 41332 MW; 29406725CF9D3512 CRC64;

Query Match 70.4%; Score 1480; DB 2; Length 381;
Best Local Similarity 72.8%; Pred. No. 1.2e-111;
Matches 267; Conservative 47; Mismatches 51; Indels 2; Gaps 2;

QY 2 LPVGNPSPSLIDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDAPKTF 61
DQ 16 LPVGNPSPSLIDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDAPKTF 75
QY 62 SWGAKPTG-SAAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVCTLGASN 120
DQ 76 SWGAAPTSPDNPADSTVTKQRONPAYKGMHDAEFTNAGYFALNIWDRFVCTLGASN 135
QY 121 GYIRGNSTAFNLVGLFGVKTNNANLNPVLSNGVVELYTDTSFWSVGARGALWECG 180
DQ 136 GYFKGNSSFNILGIGISSSTLN-DKLPNANISNGVVELYTDTSFWSVGARGALWECG 194
QY 181 CATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGKTSAT 240
DQ 195 CATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGKTSAT 254
QY 241 INYHEWQVGSLSYRLNLSVPIYGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLG 300
DQ 255 INYHEWQVGSLSYRLNLSVPIYGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLG 314
QY 301 NATALSTTDSFDMQIVSQINKFKSRKACGVTGATLVDAKWSLTAEARLINERAH 360
DQ 315 QATQVDTNILDALQIVSLQINKFKSRKACGVTGATLVDAKWSLTAEARLINERAH 374
QY 361 VSCQRF 367
DQ 375 LSAQCRF 381

RESULT 9
Q9ALJ5 PRELIMINARY; PRT: 390 AA.
AC Q9ALJ5;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW JERSEY 1, NJ1;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL; AF269266; AAK00247.1; -.
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 20
FT CHAIN 21 390 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 390 AA; 42042 MW; B62858403DBFA4E6 CRC64;

Query Match 70.3%; Score 1479; DB 2; Length 390;
Best Local Similarity 73.7%; Pred. No. 1.4e-111;
Matches 274; Conservative 41; Mismatches 49; Indels 8; Gaps 4;

QY 2 LPVGNPSPSLIDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDAPKTF 61
DQ 16 LPVGNPSPSLIDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDAPKTF 75
QY 62 SWGAKPTGSA--AANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVCTLGA 118
DQ 76 SWGAAPTAAAGTASCTASNTVAADRSNFAYGKHLQDAEWCTNAYALNIWDRFVCTLGA 142
QY 119 SNGYIRGNSTAFNLVGLFGVKTNNANLNPVLSNGVVELYTDTSFWSVGARGALWE 178
DQ 143 SNGYIRGNSTAFNLVGLFGVKTNNANLNPVLSNGVVELYTDTSFWSVGARGALWE 201
QY 179 CGCATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGCT 236
DQ 202 CGCATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGCT 261

Query Match 70.2%; Score 1477; DB 2; Length 392;
Best Local Similarity 74.1%; Pred. No. 2.1e-111;
Matches 275; Conservative 38; Mismatches 52; Indels 6; Gaps 4;

QY 2 LPVGNPSPSLIDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDAPKTF 61
DQ 16 LPVGNPSPSLIDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDAPKTF 75
QY 62 S-MGAKPTGSA--AANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVCTLGA 118
DQ 76 S-MGAKPTGSA--AANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVCTLGA 142
QY 119 SNGYIRGNSTAFNLVGLFGVKTNNANLNPVLSNGVVELYTDTSFWSVGARGALWE 178
DQ 143 SNGYIRGNSTAFNLVGLFGVKTNNANLNPVLSNGVVELYTDTSFWSVGARGALWE 201
QY 179 CGCATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGCT 236
DQ 202 CGCATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGCT 261

```



QY 237 KSATINYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWN 296  
 Db 262 KSATVYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLASAILNLTWN 321  
 QY 297 SLLGNATLSTDSFDMQIVSCQINKFKSRKAGCVTVGATLVADKWSLTAEARLINE 356  
 Db 322 TLLGVATTIDTSNKYADFMQIVSMQINKFKSRKAGCVTVGATLVADKWSLTAEARLINE 381  
 QY 357 RAAHVSGOFRF 367  
 Db 382 RAAHNAQFRF 392

RESULT 11  
 Q9AII1 PRELIMINARY; PRT; 380 AA.  
 ID Q9AII1  
 AC Q9AII1: 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).  
 GN OMPA.  
 OS Chlamydia pecorum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.  
 OX NCBI\_TaxID=85991;  
 RN [1]  
 RP SEQUENCE OF 40-352 FROM N.A.  
 RC STRAIN=L71;  
 RX MEDLINE=93123168; PubMed=8419295;  
 RA Kaitenboeck B., Kousoulas K.G., Storz J.;  
 RT Structures of and allelic diversity and relationships among the major  
 RT outer membrane protein (ompA) genes of the four chlamydial species.;  
 RL J. Bacteriol. 175:487-502(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=L71;  
 RX MEDLINE=21078680; PubMed=11211261;  
 RA Bush R.M., Everett K.D.;  
 RT "Molecular evolution of the Chlamydiaceae";  
 RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=L71;  
 RX Everrett K.D.E., Hamblly W.A., Andersen A.A.;  
 RA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF269280; AAK00261.1;  
 KW Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL <1 15 POTENTIAL.  
 FT CHAIN 16 380 MAJOR OUTER MEMBRANE PROTEIN.  
 SQ SEQUENCE 380 AA; 41047 MW; 829A18D3C5A85008 CRC64;

Query Match 69.9%; Score 1469.5; DB 2; Length 380;  
 Best Local Similarity 71.4%; Pred. No. 8.1e-111;  
 Matches 262; Conservative 54; Mismatches 48; Indels 3; Gaps 3;  
 QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 Db 16 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 75  
 QY 62 SMGAKPTG-SAAANYTTAVDRPNPAYNKLHDAEFTNAGFTIALNIWDRFDVFCITGASN 120  
 Db 76 LMGTAFTSPNSAASNTTAAERANPAYGKMHDAEFTNAGFTIALNIWDRFDVFCITGATS 135  
 QY 121 GIYRINSTAFNLVGLFGYKGTIVNANLPVSLNSGVVELYTDTSFSSVSGARGALWECG 180  
 Db 136 GYFKNSASFNLIIGLIGSGTLD-QKYPNASISNGVVELYTDTSFSSVSGARGALWECG 194  
 QY 181 CATLGAEFQYQAKSRVQELNVLNSVAQFTVHKPGYQGSPLPTNAGTSNASDLKKNAT 240  
 Db 195 CATLGAEFQYQAKSRVQELNVLNSVAQFTVHKPGYQGSPLPTNAGTSNASDLKKNAT 254

QY 241 INYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLG 300  
 Db 255 INYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLASPIFNLTWNPTLLG 314  
 QY 301 NATALSTDSFDMQIVSCQINKFKSRKAGCVTVGATLVADKWSLTAEARLINEARAH 360  
 Db 315 QATSVG-SDKYADTIQIVSLQINKFKSRKAGCVSMGATLLDADKWSLTAEARLINEARAH 373  
 QY 361 VSGOFRF 367  
 Db 374 LSAQCRF 380

RESULT 12  
 Q9AIJ2 PRELIMINARY; PRT; 391 AA.  
 ID Q9AIJ2  
 AC Q9AIJ2: 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).  
 GN OMPA.  
 OS Chlamydia psittaci (Chlamydia phila psittaci).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.  
 OX NCBI\_TaxID=83554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WC;  
 RX MEDLINE=21078680; PubMed=11211261;  
 RA Bush R.M., Everett K.D.;  
 RT "Molecular evolution of the Chlamydiaceae";  
 RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).  
 DR EMBL: AF269269; AAK00250.1;  
 KW Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL <1 17 POTENTIAL.  
 FT CHAIN 18 391 MAJOR OUTER MEMBRANE PROTEIN.  
 SQ SEQUENCE 391 AA; 42056 MW; B58A73DA4980408BF CRC64;

Query Match 67.7%; Score 1423; DB 2; Length 391;  
 Best Local Similarity 71.1%; Pred. No. 4.8e-107;  
 Matches 266; Conservative 44; Mismatches 56; Indels 8; Gaps 5;  
 QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 Db 18 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 77  
 QY 62 S-MGAKPTGSA--AANYTTAVDRPNPAYNKLHDAEFTNAGFTIALNIWDRFDVFCITG 118  
 Db 78 SGMAIPTESSGTSSAKOAVDRVNLAYGKHLQDAEFTNAGFTIALNIWDRFDVFCITG 137  
 QY 119 SNGYRINSTAFNLVGLFGYKGTI-VNA--NELPNVSLNSGVVELYTDTSFSSVSGARGA 175  
 Db 138 SNGYFKGSSAFLNLVGLFGIAGNESNALDOLPNVAITQGIIVEFYTDTSFSSVSGARGA 197  
 QY 176 LWECGATLGAEFQYQAKSRVQELNVLNSVQSVNPKYKGVVA--FPLPTDAGVATA 233  
 Db 198 LWECGATLGAEFQYQAKSRVQELNVLNSVQSVNPKYKGTI--FPLPTDAGVATA 257  
 QY 234 TGTSKATNYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNITA 293  
 Db 258 TDTKSATIKYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNITA 317  
 QY 294 WNPSSLGNATLSTDSFDMQIVSCQINKFKSRKAGCVTVGATLVADKWSLTAEARL 353  
 Db 318 WNPSSLGNATLSTDSFDMQIVSCQINKFKSRKAGCVTVGATLVADKWSLTAEARL 377  
 QY 354 INERAHVSGOFRF 367  
 Db 378 INERAHVSGOFRF 391



RX	MEDLINE=94103736; PubMed=8277245;
RA	Storey C., Lusher M., Yates P., Richmond S.;
RT	"Evidence for Chlamydia pneumoniae of non-human origin.";
RL	J. Gen. Microbiol. 139:2621-2626(1993).
DR	EMBL; I04980; AAL17396.1;
DR	InterPro: IPR000604; Chlamydia_OMP.
DR	Pfam: PF01308; Chlamydia_OMP; I.
DR	PRINTS; PR01334; CHLAMIDIAMP.
DR	ProDom; PD001717; Chlamydia_OMP; 1.
KW	Signal.
FT	SIGNAL
FT	CHAIN
SQ	SEQUENCE
Query Match	67.5%; Score 1419; DB 2; Length 402;
Best Local Similarity	70.5%; Pred. No. 1.1e-106;
Matches	268; Conservative 37; Mismatches 61; Indels 14; Gaps
QY	2 LPVGNPSDLLIDGTIWEGAGDCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
DB	23 LPVGNPAEPSLLIDGTMWEGASDCPCATWCDAISIRAGYVG DYVDFRILKVDVNKT 82
QY	62 S-MGARP---TGSA--ANYTTAVDRNPAYNKHLDHAEFTNAGFIALNIWDRFVFCTL 116
DB	83 SGMAATPTQATGNASTNOPEANGRENIAYGRHQDAEWFSAFLALINIDRFDEICTL 142
QY	117 GASNGYIRGNSAFNLVLGLGVKGTVTNANE----LPNVSLSGNVVELYDTDSFSWSVGA 172
DB	143 GASNGYFKSSAAFNLVGLIGFSATSSTETELPMQLPNVGITQGIVVEFYDTDSFSWSVGA 202
QY	173 RGALWECCGATLGAEFOYAQSPPKVBEELNVICNVSFYNKKPKYGVA--PPLPTDAGV 230
DB	203 RGALWECCGATLGAEFOYAQSPPKIEVLNVTSPPAQFVIHKPRGYKGSNFPLPTAGT 262
QY	231 AFATGKTSATINHEWQVGSLSYRLNSLVPYIGVOWSRATEDADNIRIAQPKLPFAVLN 290
DB	263 TEATDTKSATIKYHEWQVGLALSRYLNMLVPIYGVNWSRATEDADTIRIAQPKLSEILN 322
QY	291 LTAWNPSLLGNATALSTT---DSFSDFMQIVSQINKFKSRKACGVTVGATLVLDADKWSL 347
DB	323 ITWNPSLLGSTALTENNAGKDVLSDVLQIASIQINKMSRKACGVAVGATLIDADKWSI 382
QY	348 TAEARLINERAHVSGQFR 367
DB	383 TGEARLINERAHMNAQFR 402
RESULT 15	
ID	Q9AIJ9 PRELIMINARY; PRT; 382 AA.
AC	Q9AIJ9; 01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN	OMPA.
OS	Chlamydia psittaci (Chlamydophila psittaci).
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX	NCBI_TaxID=83554;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=MENINGOPNEUMONITIS, MN, ATCC VR122;
RX	MEDLINE=21078680; PubMed=11211261;
RA	Bush R.M., Everett K.D.;
RT	"Molecular evolution of the Chlamydiaceae.";
RL	Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR	EMBL; AF269262; AAK00243.1;
KW	Signal.
FT	NON_TER
FT	SIGNAL
FT	CHAIN
SQ	SEQUENCE

outer membrane protein (ompA) genes of the four chlamydial species.;  
 J. Bacteriol. 175:487-502(1993).  
 EMBL; M73040; RAD29103.1;  
 InterPro: IPR000604; Chlamydia\_OMP.  
 Pfam: PF01308; Chlamydia\_OMP; 1.  
 ProDom: PD001717; Chlamydia\_OMP; 1.  
 FT NON\_TER 1  
 FT NON\_TER 341 341  
 SQ SEQUENCE 341 AA; 36762 MW; B5933C9BF6AAF171 CRC64;

Query Match 67.5%; Score 1420.5; DB 2; Length 341;  
 Best Local Similarity 75.4%; Pred. No. 6.4e-107;  
 Matches 257; Conservative 35; Mismatches 48; Indels 1; Gaps 1;

QY 16 GTIWEAGDPCDCATWCDAISLRAGFYGDYVDFRILKVDAPKTF-SMGAKPTGSAAN 74  
 DB 1 GTMEGASGDCPCSTWCDAISIRAGYVG DYVDFRILKVDVNKTITGMCAVPTGTAAAN 60  
 QY 75 YTTAVDRNPAYNKHLDHAEFTNAGFIALNIWDRFVFCTLGASNGYIRGNSAFNLVG 134  
 DB 61 YKPTDRPNIAYGKHLQDTEWFTNAFLALNIWDRFDEICTLGASNGYFKASSAANLVG 120  
 QY 135 LFVGKGTIVANELLPNVSLSGNVVELYDTDSFSWSVCARGALWECCGATLGAEFOYAQS 194  
 DB 121 LGVGGSSIAADQPNVGITQGIVVEFYDTDSFSWSVCARGALWECCGATLGAEFOYAQN 180  
 QY 195 PKVEELNVICNVSFYNKKPKYGVAFFLTAGVATATGTSATINHEWQVGSLSY 254  
 DB 181 PKIELNVSSPAQFVVKPRGYKTAFLPLTAGTDQATDTKSAIKYHEWQVGLALS 240  
 QY 255 RLNSLVPYIGVOWSRATEDADNIRIAQPKLPFAVLNITANNPSLLGNATALSTDSFSD 314  
 DB 241 RLNLMLVPIYGVNWSRATEDADNIRIAQPKLAVALNLTWNPTLLGEATLDSNKFADF 300  
 QY 315 MQIVSQINKFKSRKACGVTVGATLVLDADKWSLTAEARLIN 355  
 DB 301 LQIASIQINKMSRKACGVAVGATLIDADKWSITGEARLIN 341

RESULT 14  
 ID Q46193 PRELIMINARY; PRT; 402 AA.  
 AC Q46193; 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.  
 GN MOMP.  
 OS Chlamydia psittaci (Chlamydophila psittaci).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 OX NCBI\_TaxID=83554;  
 RN [1]  
 RP SEQUENCE FROM N.A.







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:34:40 ; Search time 137.02 seconds  
(without alignments)  
198.401 Million cell updates/sec

Title: US-09-391-606-16

Perfect score: 1952

Sequence: 1 MLPVGNPDPSSLIDGTWE.....TAEARLINERAAHVSGQFRF 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1925	98.6	391	20	AAV35319
2	1647	84.4	343	20	AAV56771
3	1534.5	78.6	389	20	AAW98188
4	1408	72.1	402	20	AAW98189
5	1408	72.1	525	21	AAW13645
6	1408	72.1	525	22	AAW983213
7	1407	72.1	402	20	AAW98187
8	1356.5	69.5	387	20	AAV56767
9	1349	69.1	404	20	AAV56768
10	1345.5	68.9	376	19	AAW76366
11	1338.5	68.6	397	20	AAV56766

12	1336.5	68.5	397	20	AAV56763
13	1326	67.9	396	20	AAV56764
14	1321.5	67.7	397	20	AAV56765
15	1320.5	67.6	376	19	AAW76363
16	1313.5	67.3	393	20	AAV56757
17	1312.5	67.2	393	20	AAV56759
18	1308.5	67.0	372	19	AAW76365
19	1308.5	67.0	393	22	AAE06646
20	1306.5	66.9	374	19	AAW76364
21	1305.5	66.9	395	20	AAV56762
22	1300	66.6	373	19	AAW76362
23	1295	66.3	394	18	AAW15149
24	1295	66.3	394	19	AAW73141
25	1295	66.3	394	21	AAW81268
26	1294	66.3	394	19	AAW57775
27	1293	66.2	394	20	AAV56761
28	1289	66.0	392	20	AAV56760
29	1287	65.9	394	20	AAV56758
30	1284	65.8	394	20	AAV37494
31	1280	65.6	394	7	AAV60004
32	1264.5	64.8	343	20	AAV56769
33	1248	63.9	457	21	AAV82388
34	1214.5	62.2	484	21	AAV82389
35	1182	60.6	356	20	AAV56770
36	1134.5	58.1	514	21	AAV82391
37	1085.5	55.6	454	21	AAV82390
38	1053	53.9	349	21	AAV82392
39	913	46.8	228	20	AAW98184
40	907	46.5	225	20	AAW98186
41	842.5	43.2	222	20	AAW98183
42	825.5	42.3	215	20	AAW98185
43	767	39.3	277	21	AAV82393
44	519	26.6	100	20	AAW95307
45	489	25.1	100	20	AAW95295

## ALIGNMENTS

RESULT 1

AAV35319  
ID AAV35319 standard; Protein; 391 AA.

AC AAV35319;

DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae transmembrane protein sequence.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN WO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

(GEST ) GENSET.

PI Griffais R;

DR WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

XX Page 1130-1131; Disclosure; 1912pp; English.

PS



```

XX AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX Sequence 391 AA;

Query Match 98.6%; Score 1925; DB 20; Length 391;
Best Local Similarity 98.9%; Pred. No. 2e-182; Indels 0; Gaps 0;
Matches 362; Conservative 0; Mismatches 4;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGYVDFDRILKVDAPKTF 61
DB 26 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGYVDFDRILKVDAPKTF 85

QY 62 SMGAKPTGSAANYTAVDRPNPYNKHLHDAEFTNAGFTALNIWDRFDVFTLGASNG 121
DB 86 SMGAKPTGSAANYTAVDRPNPYNKHLHDAEFTNAGFTALNIWDRFDVFTLGASNG 145

QY 122 YTRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVELYDTTSFWSVGARGALWECGC 181
DB 146 YTRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVELYDTTSFWSVGARGALWECGC 205

QY 182 ATLGAEFOYAOQSKPKVEELNVICNVSQFVNPKPKYKGVAFPLPTDAGVATATGKTSATI 241
DB 206 ATLGAEFOYAOQSKPKVEELNVICNVSQFVNPKPKYKGVAFPLPTDAGVATATGKTSATI 265

QY 242 NYHEWQVGASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 301
DB 266 NYHEWQVGASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 325

QY 302 ATALSTDSFDMQIVSCQINKFKSRKACGVTGATLVADADKWSLTAEARLINERAAHV 361
DB 326 ATALSTDSFDMQIVSCQINKFKSRKACGVTGATLVADADKWSLTAEARLINERAAHV 385

QY 362 SQGFRF 367
DB 386 SQGFRF 391

RESULT 2
AAY56771
ID AAY56771 standard; Protein; 343 AA.
XX
AC AAY56771;
XX
DT 22-FEB-2000 (first entry)
XX
DE C. trachomatis serovar HuPn MOMP sequence.
XX
KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
XX
OS Chlamydia trachomatis.
XX
PN W09951745-A2.
XX
PD 14-OCT-1999.
XX
XX
PF 07-APR-1999; 99WO-CA00292.
XX
PR 07-APR-1998; 98US-0055765.
XX
PA (UYMA-) UNIV MANITOBA.

```

```

XX Bruhnam RC;
XX WPI; 1999-620205/53.
XX Non-replicating vector encoding fragments of the outer membrane protein
XX of Chlamydia, useful in vaccines and as immunogen
XX Disclosure; Fig 10 A-F; 52pp; English.
XX
CC The invention provides a non-replicating vector that comprises, linked
CC to a promoter, a nucleotide sequence that encodes a region containing at
CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
CC vaccines to generate a protective immune response (mainly cellular)
CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
CC in standard immunoassays. Immunization with the vector induces a broad
CC spectrum of immune responses, including Th1-like CD4 responses and
CC mucosal immunity, providing significant protection against subsequent
CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
CC of serovars of C. trachomatis.
XX Sequence 343 AA;

Query Match 84.4%; Score 1647; DB 20; Length 343;
Best Local Similarity 86.3%; Pred. No. 6.4e-155;
Matches 316; Conservative 2; Mismatches 2; Indels 46; Gaps 1;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGYVDFDRILKVDAPKTF 61
DB 24 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGYVDFDRILKVDAPKTF 83

QY 62 SMGAKPTGSAANYTAVDRPNPYNKHLHDAEFTNAGFTALNIWDRFDVFTLGASNG 121
DB 84 SMGAKPTGSAANYTAVDRPNPYNKHLHDAEFTNAGFTALNIWDRFDVFTLGASNG 143

QY 122 YTRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVELYDTTSFWSVGARGALWECGC 181
DB 144 YTRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVELYDTTSFWSVGARGALWECGC 193

QY 182 ATLGAEFOYAOQSKPKVEELNVICNVSQFVNPKPKYKGVAFPLPTDAGVATATGKTSATI 241
DB 194 ATLGAEFOYAOQSKPKVEELNVICNVSQFVNPKPKYKGVAFPLPTDAGVATATGKTSATI 217

QY 242 NYHEWQVGASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 301
DB 218 NYHEWQVGASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 277

QY 302 ATALSTDSFDMQIVSCQINKFKSRKACGVTGATLVADADKWSLTAEARLINERAAHV 361
DB 278 ATALSTDSFDMQIVSCQINKFKSRKACGVTGATLVADADKWSLTAEARLINERAAHV 337

QY 362 SQGFRF 367
DB 338 NSGFRF 343

RESULT 3
AAY98188
ID AAY98188 standard; Protein; 389 AA.
XX
AC AAY98188;
XX
DT 05-JUL-1999 (first entry)
XX
DE Chlamydia psittaci major outer membrane protein.
XX
KW Major outer membrane protein; MOMP; psittacosis; infection;
KW vaccine; genetic immunisation.
XX
OS Chlamydia psittaci.
XX

```



PN W09910005-A1.  
 XX 04-MAR-1999.  
 PD 28-AUG-1998; 98WO-US17943.  
 XX 28-AUG-1997; 97US-0057147.  
 XX (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.  
 PA Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
 PI WPI: 1999-254214/21.  
 XX N-PSDB; AAX25047.  
 XX A new vaccine for Chlamydia psittaci infections  
 PT Disclosure; Page 60-61; 72pp; English.  
 PS  
 XX The present sequence is the major outer membrane protein (MOMP)  
 CC of Chlamydia psittaci strain B577. A claimed MOMP polypeptide (see  
 CC AAW98184) comprises regions VD3 and VD4 of B577 MOMP, i.e. it lacks  
 CC regions VD1 and VD2. A claimed vaccine composition includes MOMP  
 CC polypeptide lacking VD1 and VD2, optionally fused to a maltose  
 CC binding protein. Also claimed are an isolated nucleic acid  
 CC encoding the polypeptide, a vector, and a method of preventing C.  
 CC psittaci infection by administering the vaccine containing the  
 CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking  
 CC regions VD1 and VD2 are useful for genetic vaccination. The  
 CC vaccines are used to prevent C. psittaci infection, especially in  
 CC birds.  
 XX  
 SQ Sequence 389 AA;  
 Query Match 78.6%; Score 1534.5; DB 20; Length 389;  
 Best Local Similarity 75.7%; Pred. No. 1.1e-143;  
 Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;  
 QY 2 LPVGNPSPSLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 DB 23 LPVGNPAEPEPSLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDVNTKI 82  
 QY 62 S-MGAKPTGSAANTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVDFVCFPLGASN 120  
 DB 83 TGMGAVPTGTAANKYKTPDRPNLAYGKHLGDAEFTNAGFTALNIWDRFDFICTLIGASN 142  
 QY 121 GYIRGNSTAFNLVGLFGVKGTTVNANLNPVSLNSGVVELYTDTSFSWSVGARGALWECG 180  
 DB 143 GYFKASSAANFLVGLFGVKGSSIAADQIPNVGITGGIVGFYDTCFWSVGARGALWECG 202  
 QY 181 CATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYGVAFPLPTDAGVATATGKTSAT 240  
 DB 203 CATLGAEFQYAGSNPKIENLVNVSQAQFVHVHPRGYGTAFLPTAGTQGTATKTSAT 262  
 QY 241 INYHEMVCASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNTAWNPSLLG 300  
 DB 263 IKYHEWQVGLSRYRLNLMPYISVNVSRATFDADAIRIAQPKLAAVINNTWPTLLG 322  
 QY 301 NATALSTTDSFDMQIVSCQINPKFSKACGVTVGATLVADADKWSLTAEARLINERAH 360  
 DB 323 EATALDTSNKTFADFIQIASIQINKMKSRKACGVAVGATLIDADKWSITGEARLINERAH 382  
 QY 361 VSGQRF 367  
 DB 383 mnaqfrf 389

RESULT 4  
 AAW98189  
 ID AAW98189 standard; Protein: 402 AA.  
 XX  
 AC AAW98189;

XX 05-JUL-1999 (first entry)  
 DT Chlamydia psittaci major outer membrane protein.  
 XX Major outer membrane protein; MOMP; psittacosis; infection;  
 KW vaccine; genetic immunisation.  
 XX Chlamydia psittaci.  
 OS W09910005-A1.  
 XX PN 04-MAR-1999.  
 XX PD 28-AUG-1998; 98WO-US17943.  
 XX PF 28-AUG-1997; 97US-0057147.  
 XX PR (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.  
 XX PA Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
 XX PI WPI: 1999-254214/21.  
 XX DR N-PSDB; AAX25048.  
 XX A new vaccine for Chlamydia psittaci infections  
 PT Disclosure; Page 65-66; 72pp; English.  
 PS  
 XX The present sequence is the major outer membrane protein (MOMP)  
 CC of Chlamydia psittaci strain 6BC. Claimed MOMP polypeptides (see  
 CC AAW98183 and AAW98184) comprise regions VD3 and VD4 of an MOMP, i.e.  
 CC they lack regions VD1 and VD2. Claimed vaccine compositions  
 CC include such MOMP polypeptides, optionally fused to a maltose  
 CC binding protein. Also claimed are isolated nucleic acids encoding  
 CC the polypeptide, a vector, and a method of preventing C. psittaci  
 CC infection by administering the vaccine containing the MOMP  
 CC polypeptide. Vectors encoding MOMP polypeptides lacking regions  
 CC VD1 and VD2 are useful for genetic vaccination. The vaccines are  
 CC used to prevent C. psittaci infection, especially in birds.  
 XX  
 SQ Sequence 402 AA;

Query Match 72.1%; Score 1408; DB 20; Length 402;  
 Best Local Similarity 70.3%; Pred. No. 4.2e-131;  
 Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;  
 QY 2 LPVGNPSPSLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 DB 23 LPVGNPAEPEPSLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDVNTKI 82  
 QY 62 S-MGAKP---TGSA--ANYTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVDFVCF 116  
 DB 83 SGMATPTQATGNASNTNQPEANGRPNIAYGRHMGDAEFNAAFILNIWDRFDFICTL 142  
 QY 117 GASNGYIRGNSTAFNLVGLFGVKG---TTVNANLNPVSLNSGVVELYTDTSFSWSVGA 172  
 DB 143 GASNGYFKASSAANFLVGLFGVKGSSIAADQIPNVGITGGIVGFYDTCFWSVGA 202  
 QY 173 RGALEWCCCATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYGVAFPLPTDAGV 230  
 DB 203 RGALEWCCCATLGAEFQYAGSNPKIENLVNVSQAQFVHVHPRGYGTAFLPTAGTQGTAT 262  
 QY 231 ATATGKTSATINYHEWQVGLSRYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290  
 DB 263 TEATDTSKTSATIKYHEWQVGLSRYRLNLMPYISVNVSRATFDADAIRIAQPKLSEIN 322  
 QY 291 LTAWNPSLLGNATAL---STTDSFDMQIVSCQINPKFSKACGVTVGATLVADADKWSL 347  
 DB 323 ITTWNPSLLIGSTALPNNSGKDVISDVLIQIASIQINKMKSRKACGVAVGATLIDADKWSI 382  
 QY 348 TAEARLINERAHVSGQRF 367







Query Match 72.1%; Score 1408; DB 22; Length 525;  
Best Local Similarity 70.3%; Pred. No. 6.3e-131;  
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSDPSLLIDGTWIEGAAGDCPCATWCDALSLRAGFYGDYVDFRILKVDAPKTF 61  
DB 146 LPVGNPAEPSLLIDGTWIEGASGDCPCATWCDALSLRAGFYGDYVDFRILKVDVNTKF 205  
QY 62 S-MGAKPT---GSAA--ANYTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVFCFL 116  
DB 206 SGMATPTQATGASNTNGPEANGRPNIAYGRHMDAEWFSNAFLAIWDRFIDYCTL 265  
QY 117 GASNGYIRGNSTAFNLVGLFGVKG----TTVNANELPNVSLNGVVELYDTDSFNSVGA 172  
DB 266 GASNGYFKASSAFLNLVGLFGVKG----TTVNANELPNVSLNGVVELYDTDSFNSVGA 325  
QY 173 RGALWECGCATIGAEFOYQAQSKPKVEELNVICNVSOFSVKNPKYKGVGA--PPLPTDAGV 230  
DB 326 RGALWECGCATIGAEFOYQAQSKPKVEELNVICNVSOFSVKNPKYKGVGA--PPLPTDAGV 385  
QY 231 ATATGKTSATINYHEWQVGSLSYRLNSLPVYIGVQWSRATEDADNIRIAQPKLPTAVLN 290  
DB 386 TEATDKSATIKYHEWQVGLAISYRLNMLVPYIGVQWSRATEDADNIRIAQPKLSEIN 445  
QY 291 LTAWNPSLLGNATAL---STTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADADKWSL 347  
DB 446 ITTWNPSLLIGSTALPNNSGKDLVLSQIAGIQLINKMKSRKACGVTVGATLVADADKWSL 505  
QY 348 TAEARLINERAAHVSGQFRF 367  
DB 506 TGEARLINERAAHMAQAFF 525

## RESULT 7

AAW98187 ID AAW98187 standard; Protein; 402 AA.

AC AAW98187;

DT 05-JUL-1999 (first entry)

DE Chlamydia psittaci major outer membrane protein.

KW Major outer membrane protein; MOMP; psittacosis; infection;

KW vaccine; genetic immunisation.

OS Chlamydia psittaci.

PN WO9910005-A1.

PD 04-MAR-1999.

PF 28-AUG-1998; 98WO-US17943.

PR 28-AUG-1997; 97US-0057147.

PA (LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.

PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;

DR WPI; 1999-254214/21.

DR N-PSDB; AAX25046.

XX A new vaccine for Chlamydia psittaci infections

PT Disclosure; Page 56-57; 72pp; English.

XX The present sequence is the major outer membrane protein (MOMP)  
CC of Chlamydia psittaci strain LSUWTK, a cocktail isolate (the  
CC MOMP gene sequence of this isolate is identical to that of C.  
CC psittaci Avian Type C). A claimed MOMP polypeptide (see AAW98183)  
CC comprises regions VD3 and VD4 of LSUWTK MOMP, i.e. it lacks  
CC regions VD1 and VD2. A claimed vaccine composition includes MOMP

CC polypeptide lacking VD1 and VD2, optionally fused to a maltose  
CC binding protein. Also claimed are an isolated nucleic acid  
CC encoding the polypeptide, a vector, and a method of preventing C.  
CC psittaci infection by administering the vaccine containing the  
CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking the  
CC regions VD1 and VD2 are useful for genetic vaccination. The  
CC vaccines are used to prevent C. psittaci infection, especially in  
CC birds.

SQ Sequence 402 AA;

Query Match 72.1%; Score 1407; DB 20; Length 402;

Best Local Similarity 70.0%; Pred. No. 5.3e-131;

Matches 266; Conservative 39; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSDPSLLIDGTWIEGAAGDCPCATWCDALSLRAGFYGDYVDFRILKVDAPKTF 61  
DB 23 LPVGNPAEPSLLIDGTWIEGASGDCPCATWCDALSLRAGFYGDYVDFRILKVDVNTKF 82  
QY 62 S-MGAKP---TGSA--ANYTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVFCFL 116  
DB 83 SGMATPTQATGASNTNGPEANGRPNIAYGRHMDAEWFSNAFLAIWDRFIDYCTL 142  
QY 117 GASNGYIRGNSTAFNLVGLFGVKG----TTVNANELPNVSLNGVVELYDTDSFNSVGA 172  
DB 143 GASNGYFKASSAFLNLVGLFGVKG----TTVNANELPNVSLNGVVELYDTDSFNSVGA 202  
QY 173 RGALWECGCATIGAEFOYQAQSKPKVEELNVICNVSOFSVKNPKYKGVGA--PPLPTDAGV 230  
DB 203 RGALWECGCATIGAEFOYQAQSKPKVEELNVICNVSOFSVKNPKYKGVGA--PPLPTDAGV 262  
QY 231 ATATGKTSATINYHEWQVGSLSYRLNSLPVYIGVQWSRATEDADNIRIAQPKLPTAVLN 290  
DB 263 TEATDKSATIKYHEWQVGLAISYRLNMLVPYIGVQWSRATEDADNIRIAQPKLSEIN 322  
QY 291 LTAWNPSLLGNATAL---STTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADADKWSL 347  
DB 323 ITTWNPSLLIGSTALPNNSGKDLVLSQIAGIQLINKMKSRKACGVTVGATLVADADKWSL 382  
QY 348 TAEARLINERAAHVSGQFRF 367  
DB 383 TGEARLINERAAHMAQAFF 402

## RESULT 8

AAW56767

ID AAW56767 standard; Protein; 387 AA.

AC AAW56767;

DT 22-FEB-2000 (first entry)

DE C. trachomatis serovar MoPh MOMP sequence.

XX Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;  
KW Cellular response; Immunogen; Th1-like CD4 response; mucosal immunity.

OS Chlamydia trachomatis.

PN WO9951745-A2.

PD 14-OCT-1999.

PF 07-APR-1999; 99WO-CA00292.

PR 07-APR-1998; 98US-0055765.

XX (UYMA-) UNIV MANITOBA.

XX Bruhnam RC;

DR WPI; 1999-620205/53.







KW Major outer membrane protein; antibody; antigen; diagnosis; disease.

OS Chlamydia trachomatis.

PN JP10234395-A.

XX 08-SEP-1998.

XX 25-FEB-1997; 97JP-0040780.

XX 25-FEB-1997; 97JP-0040780.

XX (ELED ) DENKI KAGAKU KOGYO KK.

XX WPI; 1998-535045/46.

XX Preparation of major outer membrane protein of Chlamydia trachomatis  
PT - by cloning and recombinant expression of the gene, for use as a  
PT diagnostic of Chlamydia infections

PS Claim 2; Page 15-16; 19pp; Japanese.

XX AAW76362-W76366 are major outer membrane proteins isolated from  
CC Chlamydia trachomatis which are used for the measurement of an antibody  
CC against Chlamydia trachomatis, by using it as an antigen in the form of  
CC a reagent. The method can give a diagnostic method for Chlamydia  
CC trachomatis-infected diseases with high specificity.

XX Sequence 376 AA;

Query Match 68.9%; Score 1345.5; DB 19; Length 376;

Best Local Similarity 65.2%; Pred. No. 6.1e-125;

Matches 245; Conservative 51; Mismatches 71; Indels 9; Gaps 3;

QY 1 MLPVGNPSDPSLLIDGTWEGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKT 60

DB 1 mlpvgnpaepsImidgilwegfggpcdpcatwcdaismrvgygdfvdrvktvdkne 60

QY 61 FSMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEWFTNAGFTALNIDRDFVCTL 116

DB 61 fmggaaptndaaadlqndpknvarpnaygkmgdaemfnaaymalniwdrfvactl 120

QY 117 GASNGYIRGNSTAFNLVGLFGVKGTTVANE---LPNVSLNGVVELYTDTSFWSVGAR 173

DB 121 gattgylkgnasfnlvgfktksdntaklvpnialnravveltydtttfawsvgar 180

QY 174 GALWECGCATLGAERQYAOQSKPKVEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATA 233

DB 181 aalwecgcatalgasfgyaqskpkveelnvncnaseftinkpkgygvaefpdltagtea 240

QY 234 TGTKSATINYHEWQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVINLTA 293

DB 241 tgtkdasidyhewqaslsalysrlnmftpylgvkwsvsfadrtiragpklaeaildvtt 300

QY 294 WNPSSLGNAT--ALSTTDSFDMQIVSCQINKFKSRKAGCVTVGATLVLDADKWSLTAEA 351

DB 301 lnttiagktvvasgsdndladtmqivslqinkmksrkcslagvtivdadkyavtvvet 360

QY 352 RLINERAAHVSGQFRF 367

DB 361 rlideraahvnaqfrf 376

RESULT 11

ID AAY56766

XX AAY56766 standard; Protein; 397 AA.

AC AAY56766;

DT 22-FEB-2000 (first entry)

XX C. trachomatis serovar H MOMP sequence.

XX

KW

OS

XX

OS

XX

PN

XX

XX

PD

XX

XX

PF

XX

PR

XX

XX

PA

XX

DR

XX

XX

PT

XX

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response; cellular response; immunogen; Th1-like CD4 response; mucosal immunity.

Chlamydia trachomatis.

WO9951745-A2.

14-OCT-1999.

07-APR-1999; 99WO-CA00292.

07-APR-1998; 98US-0055765.

(UYMA-); UNIV MANITOBA.

Bruhnam RC;

WPI; 1999-620205/53.

Non-replicating vector encoding fragments of the outer membrane protein of Chlamydia, useful in vaccines and as immunogen

Disclosure; Fig 10 A-F; 52pp; English.

The invention provides a non-replicating vector that comprises, linked to a promoter, a nucleotide sequence that encodes a region containing at least one of the conserved domains 2, 3 and 5 of a major outer membrane protein (MOMP) of a Chlamydia strain. The vector is used: (a) in vaccines to generate a protective immune response (mainly cellular) against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful in standard immunoassays. Immunization with the vector induces a broad spectrum of immune responses, including Th1-like CD4 responses and mucosal immunity, providing significant protection against subsequent challenge. Sequences AAY56757-71 represent MOMP sequences from a variety of serovars of C. trachomatis.

Sequence 397 AA;

Query Match 68.6%; Score 1338.5; DB 20; Length 397;

Best Local Similarity 64.8%; Pred. No. 3.3e-124;

Matches 243; Conservative 52; Mismatches 71; Indels 9; Gaps 3;

QY 2 LPVGNPSDPSLLIDGTWEGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKT 61

DB 23 lpgnpaepsImidgilwegfggpcdpcatwcdaismrvgygdfvdrvktvdknef 82

QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEWFTNAGFTALNIDRDFVCTLG 117

DB 83 gmggaaptndaaadlqndpknvarpnaygkmgdaemfnaaymalniwdrfvactlg 142

QY 118 ASNGYIRGNSTAFNLVGLFGVKGTTVANE---LPNVSLNGVVELYTDTSFWSVGAR 174

DB 143 atctgylkgnasfnlvgfktksdntaklvpnialnravveltydtttfawsvgara 202

QY 175 ALWECGCATLGAERQYAOQSKPKVEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATA 234

DB 203 alwecgcatalgasfgyaqskpkveelnvncnaseftinkpkgygvaefpdltagtea 262

QY 235 TGTKSATINYHEWQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVINLTA 294

DB 263 gtkdasidyhewqaslsalysrlnmftpylgvkwsvsfadrtiragpklaeaildvttl 322

QY 295 NPSLLGNAT--ALSTTDSFDMQIVSCQINKFKSRKAGCVTVGATLVLDADKWSLTAEA 352

DB 323 nptiagktvvasgsdndladtmqivslqinkmksrkcslagvtivdadkyavtvvet 382

QY 353 LINERAAHVSGQFRF 367

DB 383 lliideraahvnaqfrf 397



RESULT 12  
 AAY56763  
 ID AAY56763 standard; Protein; 397 AA.  
 XX  
 AC AAY56763;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE C. trachomatis serovar L3 MOMP sequence.  
 XX  
 KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;  
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN WO9951745-A2.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 07-APR-1999; 99WO-CA00292.  
 XX  
 PR 07-APR-1998; 98US-0055765.  
 XX  
 PA (UYMA-) UNIV MANITOBA.  
 PI Bruham RC;  
 XX  
 DR WPI; 1999-620205/53.  
 XX  
 PT Non-replicating vector encoding fragments of the outer membrane protein  
 PT of Chlamydia, useful in vaccines and as immunogen  
 XX  
 PS Disclosure; Fig 10 A-F; 52pp; English.  
 XX  
 CC The invention provides a non-replicating vector that comprises, linked  
 CC to a promoter, a nucleotide sequence that encodes a region containing at  
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane  
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in  
 CC vaccines to generate a protective immune response (mainly cellular)  
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful  
 CC in standard immunoassays. Immunization with the vector induces a broad  
 CC spectrum of immune responses, including Th1-like CD4 responses and  
 CC mucosal immunity, providing significant protection against subsequent  
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety  
 CC of serovars of C. trachomatis.  
 XX  
 SQ Sequence 397 AA;  
 Query Match 68.5%; Score 1336.5; DB 20; Length 397;  
 Best Local Similarity 64.8%; Pred. No. 5.1e-124; Indels 9; Gaps 3;  
 Matches 243; Conservative 50; Mismatches 73;  
 QY 2 LPVGNPSPSLIDGTIWEGAAGDCPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61  
 Db 23 lpgvnpaepsimdgilwegfgdpcpcttwdaisrmrgyggfvdfrvldvntkef 82  
 QY 62 SMGAKPTGSAANY-----TTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDVCTLG 117  
 Db 83 qmgaepttsdtaglndpntnvarpnpaygkqhmqdaemftnaaymalniwdrfdvctlg 142  
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGTTVNAE---LPNVLNSGVVVELYTDTSFSWSVGARG 174  
 Db 143 attgylkgnasfnlvgifgtktqstnfntaklvpntalnqavveltydtttfawvgara 202  
 QY 175 ALWECGCATLGAFOYQAKPVEELNVCNYSQFVSNKPKYKGYGAFPLPTDAGVATAT 234  
 Db 203 alwecgcatlgsafgysqkpkveelnvldcasetinkpkgyvgaefplldtagteaat 262  
 QY 235 GYKSATINHEWQVGSALSYRLNSLPYIGVQVRSRATEDADNRIQAOKPLPTAVLNLTAW 294  
 Db 263 gtkdasidyhewqaslsalsyrlnmtfpylgkwsrvrsdadtirtiaqpklaeavldvtcl 322

QY 295 NPSLL--GNATALSTTDSFSDFMQIVSCQINKFKSRACGVTGATLVLDADKWSLTAEAR 352  
 Db 323 nptiagkgsvvsgseneladtqivslqlnkmsrkscgiavgttividadkyavtvetr 382  
 QY 353 LINERAAHVSGQFRF 367  
 Db 383 llderaahvnaqfrf 397  
 RESULT 13  
 AAY56764  
 ID AAY56764 standard; Protein; 396 AA.  
 XX  
 AC AAY56764;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE C. trachomatis serovar A MOMP sequence.  
 XX  
 KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;  
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN WO9951745-A2.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 07-APR-1999; 99WO-CA00292.  
 XX  
 PR 07-APR-1998; 98US-0055765.  
 XX  
 PA (UYMA-) UNIV MANITOBA.  
 PI Bruham RC;  
 XX  
 DR WPI; 1999-620205/53.  
 XX  
 PT Non-replicating vector encoding fragments of the outer membrane protein  
 PT of Chlamydia, useful in vaccines and as immunogen  
 XX  
 PS Disclosure; Fig 10 A-F; 52pp; English.  
 XX  
 CC The invention provides a non-replicating vector that comprises, linked  
 CC to a promoter, a nucleotide sequence that encodes a region containing at  
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane  
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in  
 CC vaccines to generate a protective immune response (mainly cellular)  
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful  
 CC in standard immunoassays. Immunization with the vector induces a broad  
 CC spectrum of immune responses, including Th1-like CD4 responses and  
 CC mucosal immunity, providing significant protection against subsequent  
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety  
 CC of serovars of C. trachomatis.  
 XX  
 SQ Sequence 396 AA;  
 Query Match 67.9%; Score 1326; DB 20; Length 396;  
 Best Local Similarity 64.7%; Pred. No. 5.6e-123; Indels 3; Gaps 3;  
 Matches 242; Conservative 48; Mismatches 76;  
 QY 2 LPVGNPSPSLIDGTIWEGAAGDCPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61  
 Db 23 lpgvnpaepsimdgilwegfgdpcpcttwdaisrmrgyggfvdfrvldvntkef 82  
 QY 62 SMGAKPTGSAANY-----TTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDVCTLG 117  
 Db 83 qmgaepttsdtaglndpntnvarpnpaygkqhmqdaemftnaaymalniwdrfdvctlg 142  
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGTTV---NANFLPNVLSLNSGVVVELYTDTSFSWSVGARG 174  
 Db 143 attgylkgnasfnlvgifgtktqsgfdtanilvpntalnqavveltydtttfawvgara 202



QY	175	ALMECCGATLGAEFQYQAQSPKVEELNVICNVSQFSVNPFPKGYKGVAFPLPTDAGVATAT	234
Db	203	alwecgcattgasfgvaqspkveelinvicnaseftinkpkgygafplditagteaat	262
QY	235	GTSKATINVHEMOVGASLSVRLNSLVPYITGVQWSRATEDADNIRIAQPKLPTAVLNLTAW	294
Db	263	gtkdasdynewgasalsyrlnmfcpvlgvksrvsfadctirlsqpklaqplvdtl	322
QY	295	NPSLLGNATALSTDS-FSDFMQIVSQINKEFSRKACGVTVGATLVLDADKWSLTAEARL	353
Db	323	nptiagkgvtvssaeneladtmqivslqlnkmksrkscgiagttvvdadkyavtielrl	382
QY	354	INERAAHVSGQFRP	367
Db	383	ideraahvnaqfrf	396
RESULT	14		
AAV56765			
AAV56765	standard; Protein; 397 AA.		
XX	AAV56765;		
XX			
DT	22-FEB-2000 (first entry)		
XX			
XX	C. trachomatis serovar C MOMP sequence.		
XX			
KW	Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;		
KW	cellular response; immunogen; Th1-like CD4 response; mucosal immunity.		
XX			
OS	Chlamydia trachomatis.		
XX			
WN	W09951745-A2.		
PN			
PD	14-OCT-1999.		
XX			
PF	07-APR-1999; 99WO-CA00292.		
XX			
PR	07-APR-1998; 98US-0055765.		
XX			
PA	(UYMA-) UNIV MANITOBA.		
XX			
PI	Bruham RC;		
XX			
DR	WPI; 1999-620205/53.		
PT			
PT	Non-replicating vector encoding fragments of the outer membrane protein		
PT	of Chlamydia, useful in vaccines and as immunogen		
XX			
PS	Disclosure; Fig 10 A-F; 52pp; English.		
XX			
CC	The invention provides a non-replicating vector that comprises, linked		
CC	to a promoter, a nucleotide sequence that encodes a region containing at		
CC	least one of the conserved domains 2, 3 and 5 of a major outer membrane		
CC	protein (MOMP) of a Chlamydia strain. The vector is used: (a) in		
CC	vaccines to generate a protective immune response (mainly cellular)		
CC	against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful		
CC	in standard immunoassays. Immunization with the vector induces a broad		
CC	spectrum of immune responses, including Th1-like CD4 responses and		
CC	mucosal immunity, providing significant protection against subsequent		
CC	challenge. Sequences AAV56757-71 represent MOMP sequences from a variety		
XX	of serovars of C. trachomatis.		
SQ	Sequence 397 AA;		

```

Query Match      67.7%  Score 1321.5  DB 20;  Length 397;
Best Local Similarity 64.0%  Pred. No. 1.6e-122;
Matches 240;  Conservative 51;  Mismatches 75;  Indels 9;  Gaps 3;

QY      2  LPVGNPSDPSLLDGTWEGAGDPCDPCATWCDASLRACFGYGDYVFDRILKVDAPKTF 61
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

Db	23	lpvgnpaepsimldgilwsgfgdgdpdcttcwdaismrvgygdfvdrvlktdvnkef	82
Qy	62	SMGAKPTGSAANY----TTAVDRPNPAYNKHLDHAEWFTNAGFTIALNIWDRDFVECTIG	117
Db	83	qmgaaptsdtsdaglqndpttnvarppnpgaygkqmqdaemfnaaymalniwdrfdvftig	142
Qy	118	ASNGYTRGNTAFNINVLGFGVKTNYANE--LPNVLSNGVVELYTTDTSSWSVGARG	174
Db	143	atgyikgnasnlvgltgktqssfnaklipntalneavvelyinttfawsvgara	202
Qy	175	ALWECGCATLGAEFOYQAQSPKVEELNVLICNVSOFSVKNPKGYGKGVAFPLPTDAGVATAT	234
Db	203	alwecgcattlgasfyagqpkveelnvlnaseftinkpkgygaefplnitagteaat	262
Qy	235	GTSATINVHEWQVGASLSYRLNSLVPYIGVQMSRATFDADNIRIAQPKLPTAVLNLTAW	294
Db	263	gtkdasidyhewgaslalsyrlnmftpyigvkwsrvsfadtiriaqpklaeaildvtl	322
Qy	295	NPSLIGNATALS--TTDSFDEMQIVSCOINKPKSKACGVTVGATLVADAKWSLTAEAR	352
Db	323	nptiaqkgsyvsagtdneladtmgivslqlnkmksrkscgiavgttviadakyavtveta	382
Qy	353	LINERAAHVSGQFRF	367
Db	383	lideraahvnaqfrf	397
RESULT 15			
AAW76363	AAW76363 standard; protein; 376 AA.		
ID	AAW76363		
XX	AC	AAW76363;	
XX	DT	03-DEC-1998 (first entry)	
XX	DE	C. trachomatis JM109 pCR33-C major outer membrane protein.	
XX	KW	Major outer membrane protein; antigen; diagnosis; disease.	
XX	OS	Chlamydia trachomatis.	
XX	PN	JP10234395-A.	
XX	PD	08-SEP-1998.	
XX	PF	25-FEB-1997; 97JP-0040780.	
XX	PR	25-FEB-1997; 97JP-0040780.	
XX	PA	(ELED ) DENKI KAGAKU KOGYO KK.	
XX	DR	WPI; 1598-535045/46.	
XX	PT	Preparation of major outer membrane protein of Chlamydia trachomatis	
XX	PS	by cloning and recombinant expression of the gene, for use as a	
XX	PP	diagnostic of Chlamydia infections	
XX	PP	Claim 2; Page 12; 19pp; Japanese.	
XX	AAW76362	AAW76362-W76366 are major outer membrane proteins isolated from	
XX	CC	Chlamydia trachomatis which are used for the measurement of an antibody	
XX	CC	against Chlamydia trachomatis, by using it as an antigen in the form	
XX	CC	a reagent. The method can give a diagnostic method for Chlamydia	
XX	CC	trachomatis-infected diseases with high specificity.	
XX	SQ	Sequence 376 AA;	

```

Query Match      67.6%; Score 1320.5; DB 19; Length 376;
Best Local Similarity 64.4%; Pred. No. 1.8e-122;
Matches 242; Conservative, 49; Mismatches 76; Indels 9; Gaps 3;
QY 1 MJLPVGNPSDLLIDGTWEGAAGDPDPCATWCDATSLRAGFYGDYVDFRLKVDAKPT 60

```



```
Db 1 mlpvgnpaepslmidgilwegfgdpcdpottwcdaismrvyygdfvdrvlktdvke 60
Qy 61 FPMGAKPTGSAANY---TTAVDRPNPAYNKHLDHDAEFTNAGFIALNIWDREDFECTL 116
Db 61 fmqdaapttsdvaglgndptinvarpnpaygkhmqdaemftnaaymalniwdrfdvictl 120
Qy 117 GASNGYIRGNSTAFNLVGLFGVKGTNVNANE--LPNVSLSGVGVVLYTDTFSFSVSGAR 173
Db 121 gattgyikgnasfmlvglfytktgssfntaklipntalneavvelyinttfawsvgar 180
Qy 174 GALWECGCATLGAEFOYAQSKPKVEELNVICNVSQFSVKNPKYKGVAFPPLPTDAGVATA 233
Db 181 aalwecgcattlgasfqyaqskpkveelnvlnaseftinkpkgyvgaefplnitagteaa 240
Qy 234 TGTKSATINYHEWQVGASLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLNLT 293
Db 241 tgtkdasidyhewqasialsyrlnmftpyigkwsrvsfdattiriaqpklaeaildvtt 300
Qy 294 WNPSSLGNATALS--TTDSFDEMQIVSCQINKFKSRKACGVTVGATLVADAKWSLTAE 351
Db 301 lnrttagkgsvvsgtdneladtmgivslqinkmksrkscgiavggttiivdadkyavtvea 360
Qy 352 RLINERAAHVSGQPRF 367
Db 361 rlideraahvnaqfrf 376
```

Search completed: February 7, 2002, 21:34:40  
Job time: 20759 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: February 7, 2002, 21:36:25 ; Search time 85.91 seconds  
(without alignments)  
96.132 Million cell updates/sec

Title: US-09-391-606-16  
Perfect score: 1952  
Sequence: 1 MLPVGNPSPDSLLIDGTWE.....TAEARLINERAAHVSGQFRF 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136.5	7.0	61	US-08-374-560-1	Sequence 1, Appli
2	125	6.4	42	PCT-US93-08739-3	Sequence 3, Appli
3	124	6.4	25	US-08-446-692-45	Sequence 45, Appl
4	124	6.4	25	US-08-488-351A-45	Sequence 45, Appl
5	124	6.4	25	PCT-US93-08739-1	Sequence 1, Appli
6	124	6.4	25	PCT-US95-13841-17	Sequence 17, Appl
7	124	6.4	37	US-08-446-692-24	Sequence 24, Appl
8	124	6.4	37	US-08-488-351A-24	Sequence 24, Appl
9	119	6.1	31	US-08-513-365B-1	Sequence 1, Appli
10	115	5.9	25	US-09-100-409A-51	Sequence 51, Appl
11	101.5	5.2	919	US-08-788-674-4	Sequence 4, Appli
12	97.5	5.0	457	US-09-142-759-1	Sequence 1, Appli
13	90.5	4.6	784	US-09-371-913A-7	Sequence 7, Appli
14	87.5	4.5	834	US-08-471-033-21	Sequence 21, Appl
15	87.5	4.5	834	US-08-471-044-21	Sequence 21, Appl
16	87.5	4.5	834	US-08-463-483A-21	Sequence 21, Appl
17	87.5	4.5	834	US-08-471-046A-21	Sequence 21, Appl
18	87.5	4.5	834	US-08-470-566B-21	Sequence 21, Appl
19	87.5	4.5	834	US-08-469-334-21	Sequence 21, Appl
20	87.5	4.5	834	US-09-300-529-21	Sequence 21, Appl
21	85	4.4	621	US-08-262-220-8	Sequence 8, Appli
22	85	4.4	621	US-08-471-733-8	Sequence 8, Appli
23	85	4.4	621	US-08-468-878-8	Sequence 8, Appli
24	85	4.4	621	US-08-750-494-8	Sequence 8, Appli
25	85	4.4	628	US-07-952-853-22	Sequence 22, Appl
26	85	4.4	628	US-08-914-848-22	Sequence 22, Appl
27	85	4.4	664	US-09-295-186-17	Sequence 17, Appl

28	81	4.1	306	1	US-08-312-387B-4	Sequence 4, Appli
29	81	4.1	306	1	US-08-683-426-4	Sequence 4, Appli
30	81	4.1	306	1	US-08-683-458-4	Sequence 4, Appli
31	81	4.1	306	2	US-08-878-360-4	Sequence 4, Appli
32	81	4.1	306	3	US-08-478-140B-4	Sequence 4, Appli
33	81	4.1	400	1	US-08-118-906-14	Sequence 14, Appl
34	81	4.1	400	1	US-08-486-196-14	Sequence 14, Appl
35	81	4.1	400	1	US-08-488-135-14	Sequence 14, Appl
36	81	4.1	400	2	US-08-474-063-14	Sequence 14, Appl
37	81	4.1	400	4	US-09-233-506-4	Sequence 4, Appli
38	81	4.1	592	1	US-08-217-327-8	Sequence 8, Appli
39	81	4.1	1338	2	US-08-728-470-9	Sequence 9, Appli
40	81	4.1	1338	4	US-08-719-641-9	Sequence 9, Appli
41	81	4.1	1529	2	US-08-728-470-10	Sequence 10, Appl
42	81	4.1	1529	4	US-08-719-641-10	Sequence 10, Appl
43	81	4.1	1599	2	US-08-617-697-9	Sequence 9, Appli
44	81	4.1	1600	2	US-08-617-697-10	Sequence 10, Appl
45	80.5	4.1	667	1	US-08-471-033-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-08-374-560-1  
Sequence 1, Application US/08374560  
Patent No. 5882645  
GENERAL INFORMATION:  
APPLICANT: TOTH, Istvan  
APPLICANT: GIBBONS, William Anthony  
TITLE OF INVENTION: PEPTIDE COMPOUNDS  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Ostrolenk, Faber, Gerb & Soffen  
STREET: 1180 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10036-8403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/374,560  
FILING DATE: 13-MAR-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9215780.9  
FILING DATE: 24-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meilman, Edward A.  
REGISTRATION NUMBER: 24,735  
REFERENCE/DOCKET NUMBER: P/365-302  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 382-0700  
TELEFAX: (212) 382-0888  
TELEX: 236925  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 61 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-374-560-1

Query Match 7.0%; Score 136.5; DB 2; Length 61;  
Best Local Similarity 47.5%; Pred. No. 3.2e-07;  
Matches 29; Conservative 10; Mismatches 21; Indels 1; Gaps 1;



```
QY 269 RATEADNIRTAQKLPVAVLNLTAWNPISLGNATALSTTD-SFSDFMQIVSCQINKFKS 327
      ||||||| ||||||| : :|| |||: : : : ||||| :||| ||
Db 1 RASFDADTIRIAQPKSAETIFDVTTLPTIAGAGDVKTSAEGQLGDTMQIVSLQLNKWK 60
      : :|| |||: : :|| |||: : : : ||||| :||| ||
QY 328 R 328
      |
Db 61 R 61

RESULT 2
PCT-US93-08739-3
; Sequence 3, Application PC/TUS9308739
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented
; APPLICANT: by the Secretary of the Department of Health and Human Services
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR
; TITLE OF INVENTION: CHLAMYDIA
; TITLE OF INVENTION: TRACHOMATIS
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08739
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/947,671 US
; FILING DATE: 18 SEP 93
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
PCT-US93-08739-3

Query Match 6.4%; Score 125; DB 5; Length 42;
Best Local Similarity 61.9%; Pred. No. 2.8e-06;
Matches 26; Conservative 4; Mismatches 2; Indels 10; Gaps 2;

QY 103 ALNIWDRFDVFTLGASNGYIRGNSTAFNLVLGLGVKGTIVN 144
      ||||||| ||||||| : :|| |||: : : : ||||| :||| ||
Db 1 ALNIWDRFDVFTLGATGTLKGNSTAFNLVLGLGVKGTIVN 32
      : :|| |||: : :|| |||: : : : ||||| :||| ||

RESULT 3
US-08-446-692-45
; Sequence 45, Application US/08446692
; Patent No. 5759551
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053

Query Match 6.4%; Score 125; DB 1; Length 25;
Best Local Similarity 84.0%; Pred. No. 1.6e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNSTAFNLVLGLGVKGTIVN 144
      ||||||| ||||||| : :|| |||: : : : ||||| :||| ||
Db 1 ALNIWDRFDVFTLGATGTLKGNSTAFNLVLGLGVKGTIVN 32
      : :|| |||: : :|| |||: : : : ||||| :||| ||

RESULT 4
US-08-488-351A-45
; Sequence 45, Application US/08488351A
; Patent No. 5843446
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053

Query Match 6.4%; Score 124; DB 1; Length 25;
Best Local Similarity 84.0%; Pred. No. 1.6e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNSTAFNLVLGLGVKGTIVN 144
      ||||||| ||||||| : :|| |||: : : : ||||| :||| ||
Db 1 ALNIWDRFDVFTLGATGTLKGNSTAFNLVLGLGVKGTIVN 32
      : :|| |||: : :|| |||: : : : ||||| :||| ||

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,692
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maria C.H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4146 US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)415-8745
TELEFAX: (516)751-6849
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-446-692-45
```



ATTORNEY/AGENT INFORMATION:  
NAME: Maria C.H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-488-351A-45

Query Match 6.4%; Score 124; DB 2; Length 25;  
Best Local Similarity 84.0%; Pred. No. 1.6e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRFDVFTLGASNGYIRGNS 127  
|||||  
Db 1 ALNIWDRFDVFTLGATTGILKGN 25

## RESULT 5

PCT-US93-08739-1  
Sequence 1, Application PC/TUS9308739  
GENERAL INFORMATION:  
APPLICANT: The Government of the United States of America  
APPLICANT: as represented  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR  
TITLE OF INVENTION: CHLAMYDIA  
TITLE OF INVENTION: TRACHOMATIS  
NUMBER OF SEQUENCES: 3  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/08739  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/947,671 US  
FILING DATE: 18 SEP 93  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
PCT-US93-08739-1

Query Match 6.4%; Score 124; DB 5; Length 25;  
Best Local Similarity 84.0%; Pred. No. 1.6e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRFDVFTLGASNGYIRGNS 127  
|||||  
Db 1 ALNIWDRFDVFTLGATTGILKGN 25

## RESULT 6

PCT-US95-13841-17  
Sequence 17, Application PC/TUS9513841  
GENERAL INFORMATION:

APPLICANT: United Biomedical Inc; Walfield, Alan M.;  
APPLICANT: Wang, Chang Yi  
TITLE OF INVENTION: Synthetic IgE Membrane Anchor  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C.H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13841  
FILING DATE: 25-OCT-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/328,519  
FILING DATE: 25-OCT-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lin, Maria C.H.  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4117  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-13841-17

Query Match 6.4%; Score 124; DB 5; Length 25;  
Best Local Similarity 84.0%; Pred. No. 1.6e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRFDVFTLGASNGYIRGNS 127  
|||||  
Db 1 ALNIWDRFDVFTLGATTGILKGN 25

## RESULT 7

US-08-446-692-24  
Sequence 24, Application US/08446692  
Patent No. 5759551  
GENERAL INFORMATION:

APPLICANT: Ladd, Anna  
APPLICANT: Wang, Chang Yi  
APPLICANT: Zamb, Timothy  
TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C.H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10154-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS



;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/446,692  
;; FILING DATE: 7-JUN-1995  
;; CLASSIFICATION: 424  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Maria C.H. Lin  
;; REGISTRATION NUMBER: 29,323  
;; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212)415-8745  
;; TELEFAX: (516)751-6849  
;; INFORMATION FOR SEQ ID NO: 24:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 37 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-446-692-24

Query Match 6.4%; Score 124; DB 1; Length 37;  
Best Local Similarity 84.0%; Pred. No. 3e-06; 1; Indels  
Matches 21; Conservative 3; Mismatches 1; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127  
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

RESULT 8  
US-08-488-351A-24  
; Sequence 24, Application US/08488351A  
; Patent No. 5843446  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; TITLE OF INVENTION: And synthetic universal immune stimulators for vaccines  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,351A  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/446,692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/229,275  
; FILING DATE: 14-APR-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/057,166  
; FILING DATE: 27-APR-1992  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2

;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212)415-8745  
;; TELEFAX: (516)751-6849  
;; INFORMATION FOR SEQ ID NO: 24:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 37 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-488-351A-24

Query Match 6.4%; Score 124; DB 2; Length 37;  
Best Local Similarity 84.0%; Pred. No. 3e-06;  
Matches 21; Conservative 3; Mismatches 1; Indels  
0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127  
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

RESULT 9  
US-08-519-385B-1  
; Sequence 1, Application US/08519385B  
; Patent No. 6001372  
; GENERAL INFORMATION:  
; APPLICANT: DeMars, Robert I.  
; APPLICANT: Ortiz, Linette (n.m.i.)  
; TITLE OF INVENTION: Antigenic Peptides of Chlamydia trachomatis  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 411 East Wisconsin Avenue  
; CITY: Milwaukee  
; STATE: WI  
; COUNTRY: U.S.A.  
; ZIP: 53202-4497  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/519,385B  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Schwartz, Carl R.  
; REGISTRATION NUMBER: 29,437  
; REFERENCE/DOCKET NUMBER: 960296.93456  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (414) 277-5715  
; TELEFAX: (414) 271-3552  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 31 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: no  
; ANTI-SENSE: no  
; FRAGMENT TYPE: internal fragment  
; ORIGINAL SOURCE:  
; ORGANISM: Chlamydia trachomatis  
; STRAIN: Serovar E  
; INDIVIDUAL ISOLATE: MOMP Protein  
US-08-519-385B-1

Query Match 6.1%; Score 119; DB 3; Length 31;  
Best Local Similarity 64.5%; Pred. No. 7.5e-06;  
Matches 20; Conservative 5; Mismatches 6; Indels  
0; Gaps 0;



ADDRESS: Carella, Byrne, Bain,  
ADDRESS: Gilfillan, Cecchi, Stewart &  
ADDRESS: Olstein  
STREET: 6 Becker Farm Road  
CITY: Roseland

RESULT 12



## US-09-142-759-1

; Sequence 1, Application US/09142759A

; Patent No. 6127160

; GENERAL INFORMATION:

; APPLICANT: YAMANOBE, Takashi

; APPLICANT: WATANABE, Manabu

; APPLICANT: HAMAYA, Toru

; APPLICANT: SUMIDA, Naomi

; APPLICANT: AOYAGI, Kaoru

; APPLICANT: MURAKAMI, Takashi

; TITLE OF INVENTION: PROTEIN HAVING CELLULASE ACTIVITY AND PROCESS FOR

; TITLE OF INVENTION: PRODUCING THE SAME

; FILE REFERENCE: Q51673

; CURRENT APPLICATION NUMBER: US/09/142.759A

; CURRENT FILING DATE: 1998-09-14

; EARLIER APPLICATION NUMBER: PCT/JP97/00824

; EARLIER FILING DATE: 1997-03-14

; EARLIER APPLICATION NUMBER: P. HEI-8-84479

; EARLIER FILING DATE: 1996-03-14

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 457

; TYPE: PRT

; ORGANISM: Acremonium cellulolyticus

US-09-142-759-1

Query Match 5.08; Score 97.5; DB 3; Length 457;

Best Local Similarity 22.38; Pred. No. 0.097;

Matches 67; Conservative 36; Mismatches 114; Indels 83; Gaps 16;

Qy 68 TGSAAATTTAVDRN-----PAYNKLHDAEFTNAG-----FIALNIWDR 109

Db 125 TGSAAATKAEIPSFVWLTAAKVPMTGTYLANEANKAGASPPTAGIFVYVDLPDR 184

Qy 110 FDVFTLGASNG-YIRGNSTAFNLVGLFGVGTTVNANELPNV-----SLNSGVVE 159

Db 185 ---DCAAAASNGEYTVANNGVANYKAY--IDSIYAQLKAYPDVHTILIEPDSLAN----- 235

Qy 160 LYDTFSFSWSVGARGALWECGCATLGAEEFYQAQSKPKVEELNVCNVSOFVSNPKPKYKG 219

Db 236 MYTNLSHTAKAEQAQAYTEC-----VNYA-----LINLN-LANVAMYIDAGHAGWLG 281

Qy 220 VAFPLPTDAGV-----ATATGTSKATINYNHEWQVGASLSYR-----LN 257

Db 282 WSNALSPAAQLFATVYKNASAPASLRGLATNVANNVANSISSPPSYTSDSNYDEKLYIN 341

Qy 258 SLVPYIGVO-WSRATFDADNIRI-AQPKLPTAVLMLTANP--SLLGNTALSTTDSFSD 313

Db 342 ALSPLLTSGWPNHAFIMDTSRNGVQP-----TKQAWGDWCNVIGTGFQVQPTTNGD 395

## RESULT 13

US-09-371-913A-7

; Sequence 7, Application US/09371913A

; Patent No. 6297369

; GENERAL INFORMATION:

; APPLICANT: Schnepf, H. Ernest

; APPLICANT: Narva, Kenneth E.

; APPLICANT: Stockhoff, Brian A.

; APPLICANT: Finstad Lee, Stacey

; APPLICANT: Walz, Mikki

; APPLICANT: Sturgis, Blake

; TITLE OF INVENTION: Pesticidal Toxins and Genes from Bacillus laterosporus

; TITLE OF INVENTION: Strains

; FILE REFERENCE: MA-719XC2 US

; CURRENT APPLICATION NUMBER: US/09/371.913A

; CURRENT FILING DATE: 1999-08-10

; PRIOR APPLICATION NUMBER: 60/095,955

; PRIOR FILING DATE: 1998-08-10

; PRIOR APPLICATION NUMBER: 60/138,251

; PRIOR FILING DATE: 1999-06-08

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 1.1

; SEQ ID NO 7

; LENGTH: 784

; TYPE: PRT

; ORGANISM: Peptide sequence

US-09-371-913A-7

Query Match 4.6%; Score 90.5; DB 4; Length 784;

Best Local Similarity 19.38; Pred. No. 1.2;

Matches 75; Conservative 49; Mismatches 137; Indels 127; Gaps 18;

Qy 46 DYVFDRIKLVADP-----KTFSGAKPTGSAANYTTAVDR-PNPAYNKLHDAEWF--- 96

Db 155 EYQPDQIVNRDSKIFKEKLFKVDSK-----QQSHQVQLDELNRNPEFNK--KETQOQFLEK 207

Qy 97 -TNAGFTALNIWDRFDVFTLGAS-----NGVIRGNSTAFNLVGLFGVKGTT----- 142

Db 208 ASKTNLFTQNMKRDEATDGDSDIPDLWEENGYTIQNKVAVKWDSDFAAKGYTKFVSNP 267

Qy 143 -----VNANEL-----PNVSLNSGVVE 159

Db 268 FDSHTVGDPTDYDEKAARDLDLANAKETFNPLVAAPFSVNVNLEKVLSPNEDLSNS-VE 326

Qy 160 LYDTFSFSWSVGARGALWECGCATLG-----AEFYQAQSKPKVEELNVCNVSOFVSNK 213

Db 327 SHSSTNNYSYT-NTGCVSIEAGSGPLGISYGSVANYQHSETVAK-EWGTTSTGTSTQFN--- 381

Qy 214 PKGYKGVAFPLPTDAGVATATGTSKATINYNHEWQVGASLSYRLASLVYPYIGVQVSRATFD 273

Db 382 -----TASAGYLNANVRYN--NVGTGAIYEVKPTTGFV-----LD 414

Qy 274 ADNIRIAQPKLPTAVLNLTAWN--PSLIGNATALTSTTDSFS-----DFMQIVSQOI 322

Db 415 NDTVATITAKSNSTALSTSPGESYPPKKGQNGIAINTWDFNSHPITLKNKQDLDQIFNNKP 474

Qy 323 NKFSRKACGV-----TVGATLVADAKWS 346

Db 475 LMLETNQADGVYKINDTSGNIVTGTGEWN 502

## RESULT 14

US-08-471-033-21

; Sequence 21, Application US/08471033

; Patent No. 5770696

; GENERAL INFORMATION:

; APPLICANT: Warren, Gregory W

; APPLICANT: Koziel, Michael G

; APPLICANT: Mullins, Martha A

; APPLICANT: Nye, Gordon J

; APPLICANT: Carr, Brian

; APPLICANT: Desai, Nalini M

; APPLICANT: Kostichka, N. Kristy

; APPLICANT: Duck, Nicholas B

; APPLICANT: Estruch, Juan J

; TITLE OF INVENTION: NO. 5770696el Pesticidal Proteins and Strains

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

; STATE: NY

; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30B

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/471,033

; FILING DATE:



```

RESULT 15
US-08-471-044-21
; Sequence 21, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Cart, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J

```

143	QY	-----VNANEL-----PNVLSLN	151	PNVLSLN	155	PNVLSLN
262	Db	VSNPLDSHTVGDPYTDYEKAARDLDSNAKETFNPLVAAPPSVNVSMKVKVILSPNENLSN	321	PNVLSLN	321	PNVLSLN
156	QY	GVVELYTDTSFSGVSGARGALMECCGATLGAEF-----QYAAQSKPKVEELNVIC-NVSOFS	210	QYAAQSKPKVEELNVIC	210	QYAAQSKPKVEELNVIC
322	Db	S-VESHSSNTNWSVT-NTEGASIEAGCGPLGLSFGVSVTYQHSETVAQENGSTGTGNTSQFN	379	QYAAQSKPKVEELNVIC	379	QYAAQSKPKVEELNVIC
211	QY	VNKPCKYKVAPPLPTDAGVATATGKTSATINYHEQVGCASLSYRLNSLPIVIGVQMSRA	270	QYAAQSKPKVEELNVIC	270	QYAAQSKPKVEELNVIC
380	Db	-----TASAGYLNANVRYN--NVGTGAIYDVKPTTSEV-----	410	QYAAQSKPKVEELNVIC	410	QYAAQSKPKVEELNVIC
271	QY	TFEADNIRIAQPKLPTAVLNLTAWN-----PSSLGNATALSTTDSDF-MQIVSOQI	322	QYAAQSKPKVEELNVIC	322	QYAAQSKPKVEELNVIC
411	Db	---LNNNTIA---TITAKSNSTALRISPGDSYPEIGENAIATISMDFNSHPTLLNKOOF	464	QYAAQSKPKVEELNVIC	464	QYAAQSKPKVEELNVIC



OY 323 NKFKSRK 329  
I : : |  
Db 465 NQLINK 471

Search completed: February 7, 2002, 21:36:26  
Job time: 20495 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:38:10 ; Search time 96.2 Seconds  
(without alignments)  
290.604 Million cell updates/sec

Title: US-09-391-606-16  
Perfect score: 1952  
Sequence: 1 MLPVGNPSPDLLDGTWE.....TAEARLINERAAHVSGQFRF 367

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1947	99.7	389	2 A43587	major outer membra
2	1947	99.7	389	2 D86577	major outer membra
3	1910	97.8	389	2 I40864	major outer membra
4	1882	96.4	389	2 I40739	major outer membra
5	1534.5	78.6	389	1 MMCWP3	major outer membra
6	1496.5	76.7	389	2 A60109	major outer membra
7	1477	75.7	392	2 A40371	major outer membra
8	1419	72.7	402	2 I40740	major outer membra
9	1415	72.5	402	1 MMCWPM	major outer membra
10	1410	72.2	402	2 B60109	major outer membra
11	1408	72.1	402	2 A60341	major outer membra
12	1351.5	68.2	387	2 C81747	major outer membra
13	1351	69.2	404	2 I40741	major outer membra
14	1349.5	69.1	387	2 J70947	major outer membra
15	1349.5	69.1	387	2 S16034	mouse pneumonitis
16	1340.5	68.7	397	1 MMCWTH	major outer membra
17	1338.5	68.6	397	2 JF0413	major outer membra
18	1333.5	68.3	375	2 S11007	major outer membra
19	1328	68.0	374	2 S11006	major outer membra
20	1325	67.9	396	2 S12799	major outer membra
21	1315.5	67.4	393	1 MMCWTE	major outer membra
22	1315.5	67.4	397	1 MMCWTC	major outer membra
23	1313.5	67.3	393	2 S06259	major outer membra
24	1310.5	67.1	393	2 T01645	major outer membra
25	1308.5	67.0	393	2 H71484	probable major out
26	1307.5	67.0	395	1 MMCWTF	major outer membra
27	1303.5	66.8	393	2 JC1432	major outer membra
28	1295	66.3	394	2 S11012	major outer membra
29	1292	66.2	372	2 B60756	major outer membra

30	1285	65.8	372	2 S11009	major outer membra
31	1275	65.3	394	1 MMCWTB	major outer membra
32	373	19.1	158	2 A60333	outer membrane pro
33	361	18.5	156	2 B60333	outer membrane pro
34	355.5	18.2	155	2 C60333	outer membrane pro
35	215.5	11.0	136	2 F30593	outer membrane pro
36	215	11.0	135	2 C30593	outer membrane pro
37	212.5	10.9	136	2 D30593	outer membrane pro
38	206.5	10.7	136	2 B30593	outer membrane pro
39	206.5	10.6	136	2 I30587	outer membrane pro
40	206.5	10.6	136	2 E30593	outer membrane pro
41	203.5	10.4	136	2 A30593	outer membrane pro
42	189	9.7	340	2 H71479	probable outer mem
43	186.5	9.7	134	2 G30587	outer membrane pro
44	182	9.3	340	2 H81742	major outer membra
45	179.5	9.2	134	2 H30587	outer membrane pro

## ALIGNMENTS

### RESULT 1

A43587

major outer membrane protein, porin CP0051 precursor [imported] - Chlamydoiphila pneu  
N:Alternate names: MOMP  
C:Species: Chlamydoiphila pneumoniae, Chlamydia pneumoniae  
C:Date: 29-Jan-1993 #sequence revision 29-Jan-1993 #text\_change 11-May-2000  
C:Accession: A43587; A49751; A49216; G72044; F81619  
R:Perez Melgosa, M.; Kuo, C.C.; Campbell, L.A.  
Infect. Immun. 59, 2195-2199, 1991  
A:Title: Sequence analysis of the major outer membrane protein gene of Chlamydia pne  
A:Reference number: A43587; MUID:91244474  
A:Accession: A43587  
A:Molecule type: DNA  
A:Residues: 1-389 <PER>  
A:Cross-references: GB:M69230; NID:g144540; PIDN:AAA73071.1; PID:g144541  
R:Carter, M.W.; Al-Mahdawi, S.A.H.; Giles, I.G.; Trehan, J.D.; Ward, M.E.; Clarke, J. Gen. Microbiol. 137, 465-475, 1991  
A:Title: Nucleotide sequence and taxonomic value of the major outer membrane protein  
A:Reference number: A49751; MUID:91237311  
A:Accession: A49751  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-389 <CAR>  
A:Cross-references: GB:M64064; GB:M34942; NID:g144534; PIDN:AAA23143.1; PID:g144535  
A:Note: isolate IOL-207  
R:Gaydos, C.A.; Quinn, T.C.; Bobo, L.D.; Eiden, J.J.  
Infect. Immun. 60, 5319-5323, 1992  
A:Title: Similarity of Chlamydia pneumoniae strains in the variable domain IV region  
A:Reference number: A49216; MUID:93084388  
A:Accession: A49216  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 297-352 <GAV>  
A:Cross-references: GB:S50607; NID:g260972; PIDN:AAB24363.1; PID:g260973  
A:Note: sequence extracted from NCBI backbone (NCBIN:120604, NCBIIP:120605)  
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: G72044  
A:Molecule type: DNA  
A:Residues: 1-389 <ARN>  
A:Cross-references: GB:AB001652; GB:AB001363; NID:g4376997; PIDN:AAI8834.1; PID:g43  
A:Experimental source: strain CWL029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hick  
C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbu  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPh and Chlamydia pneumoniae AR3  
A:Reference number: AB1500; MUID:20150255  
A:Accession: F81619  
A>Status: preliminary  
A:Molecule type: DNA



Db	264	NYHEWQAGSLSYRLNSLVPIGVQMSRATFDADNIRIAQPKLPTAVLNLTAWNP	3233
QY	302	ATALSTTDSFDMQIVSCQINFKFSRKACGVTVGATLVDADKWSLTAEARLINEAAHV	3611
		::	
Db	324	TTTLATSDSFDMQIVSCQINFKFSRKACGVTVGATLVDADKWSLTAEARLINEAAHV	3833
QY	362	SGOFFR	367



Db 384 SQQFRF 389

RESULT 4

I40739

major outer membrane protein precursor - Chlamydomophila pneumoniae (strain equine/W16)

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

A:Variety: strain equine/W16

C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 20-Apr-2000

C:Accession: I40739

R:Storey, C.; Lusher, M.; Yates, P.; Richmond, S.

J. Gen. Microbiol. 139, 2621-2626, 1993

A:Title: Evidence for Chlamydia pneumoniae of non-human origin.

A:Reference number: I40739; MUID:94103736

A:Accession: I40739

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-389 <STO>

C:Cross-references: GB:I40982; NID:g289840; PIDN:AAAL1797.1; PID:g289841

C:Comment: On the basis of the major outer membrane protein the authors classified the

the sequence of the genome strain CWL029 and strain IOL-207. See PIR:A43587.

C:Genetics:

A:Gene: momp

C:Superfamily: Chlamydia major outer membrane protein

C:Keywords: membrane protein

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-389/Product: major outer membrane protein #status predicted <MAT>

Query Match

Best Local Similarity 96.4%; Score 1882; DB 2; Length 389;

Matches 346; Conservative 15; Mismatches 5; Indels 0; Gaps 0;

QY 2 LPVGNPDSPLLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGYVDFRILKVDAPKTF 61

Db 24 LPVGNPDSPLLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGYVDFRILKVDAPKTF 83

QY 62 SMGAKPTGSAANNTTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFVDFCTLGASNG 121

Db 84 SMGAKPTGSAANNTTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFVDFCTLGASNG 143

QY 122 YIRGNSTAFNLVGLFVGKGTNNANLNPVSLNSGVVELYTDTSFSWSVGARGALWECG 181

Db 144 YKGNSSAANLNLVGLFVGKGTNNANLNPVSLNSGVVELYTDTSFSWSVGARGALWECG 203

QY 182 ATLGAEEFYAQSAPKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGTSAT 241

Db 204 ATLGAEEFYAQSAPKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGTSAT 263

QY 242 NYHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 301

Db 264 NYHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 323

QY 302 ATALSTTDSFSDPMQIVSCQINPKSKACGVTGATLVDAKWSLTAEARLINERAHH 361

Db 324 ATAVSSDDQFSDPMQIVSCQINPKSKACGVTGATLVDAKWSLTAEARLINERAHH 383

QY 362 SQQFRF 367

Db 384 SQQFRF 389

RESULT 5

MMCW3

major outer membrane protein precursor - Chlamydomophila psittaci (strain S26/3)

C:Species: Chlamydomophila psittaci, Chlamydia psittaci

C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 31-Mar-2000

C:Accession: S08770

R:Herring, A.J.; Tan, T.W.; Baxter, S.; Inglis, N.F.; Dunbar, S.

FEMS Microbiol. Lett. 65, 153-158, 1989

A:Title: Sequence analysis of the major outer membrane protein gene of an ovine abortion

A:Reference number: S08770

A:Accession: S08770

A:Molecule type: DNA

A:Residues: 1-389 <HER>

C:Cross-references: EMBL:X51859; NID:g40600; PIDN:CAA36152.1; PID:g40601

C:Superfamily: Chlamydia major outer membrane protein

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match

Best Local Similarity 78.6%; Score 1534.5; DB 1; Length 389;

Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;

QY 2 LPVGNPDSPLLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGYVDFRILKVDAPKTF 61

Db 23 LPVGNPAEPSSLIDGTIWEAGAGDPCDPCSTWCDATSLRAGFYGYVDFRILKVDVKNKI 82

QY 62 S-MGAKPTGSAANNTTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFVDFCTLGASN 120

Db 83 TCMGAVPTGTAAANTKTPTDRPNIAIYKHLQDAEFTNAAFLALNIWDRFDFICTLGASN 142

QY 121 GYIRGNSTAFNLVGLFVGKGTNNANLNPVSLNSGVVELYTDTSFSWSVGARGALWECG 180

Db 143 GYFKASSRAFNVLGLVGVKSSIAADOLPNVGITGIVEFYDTTFSWSVGARGALWECG 202

QY 181 CATLGAEEFYAQSAPKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGTSAT 240

Db 203 CATLGAEEFYAQSAPKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGTSAT 262

QY 241 INYHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLG 300

Db 263 IYHEWQVGLALSYRLNLVPIYSVNSRATFDADNIRIAQPKLAAVNLTAWNPSSLG 322

QY 301 NATALSTTDSFSDPMQIVSCQINPKSKACGVTGATLVDAKWSLTAEARLINERAHH 360

Db 323 EATALDTSNKFADEFLQIASIQINKMKSKACGAVGATLIDADKWSITGEARLINERAHH 382

QY 361 VSGQFRF 367

Db 383 MNAQFRF 389

RESULT 6

A60109

major outer membrane protein precursor - Chlamydomophila psittaci (strain Guinea pig

C:Species: Chlamydomophila psittaci, Chlamydia psittaci

C>Date: 10-Nov-1992 #sequence\_revision 10-Nov-1992 #text\_change 31-Mar-2000

C:Accession: A60109

R:Zhang, Y.X.; Morrison, S.G.; Caldwell, H.D.; Baehr, W.

Infect. Immun. 57, 1621-1625, 1989

A:Title: Cloning and sequence analysis of the major outer membrane protein genes of

A:Reference number: A60109; MUID:89212917

A:Accession: A60109

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-389 <ZHA>

C:Superfamily: Chlamydia major outer membrane protein

C:Keywords: membrane protein

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match

Best Local Similarity 76.7%; Score 1496.5; DB 2; Length 389;

Matches 272; Conservative 42; Mismatches 51; Indels 3; Gaps 2;

QY 2 LPVGNPDSPLLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGYVDFRILKVDAPKTF 61

Db 23 LPVGNPAEPSSLIDGTIWEAGAGDPCDPCSTWCDATSLRAGFYGYVDFRILKVDVKNKI 82

QY 62 SNGAKPTGSAANNTTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFVDFCTLGASNG 121

Db 83 SNGTAPTGAADFKTVADRNRIAYKHKMQDAEFTNAAFLALNIWDRFVDFCTLGASNG 142



QY 122 YIRGNSTAFNLVGLFGVKGVTNNANLPVLSNGWVLYTDTFSWSVGARGALWECG 181  
Db 143 YLKANAANFLVGLLGVGTDLQ -GGYPNAISQGLVELYDTTFSWSVGARGALWECG 201  
QY 182 ATLGAEFQYQAQSKPKVEELNVCNVSQSVNPKPKYKGA--FPLPTDAGVATATGTXSA 239  
Db 202 ATLGAEFQYQAQSKPKVEELNVCNVSQSVNPKPKYKGA--FPLPTDAGVATATGTXSA 261  
QY 240 TTYHEWQVGLSLVRLNLSVYICVQWNRATFDADNIRIAQPKLPTAVLNLTANNPSIL 299  
Db 262 TTYHEWQVGLSLVRLNLSVYICVQWNRATFDADNIRIAQPKLPTAVLNLTANNPSIL 321  
QY 300 GNATALSTTDSFDMQIVSCQINKFKSKAGCVTVGATLVADAKWSLTAEARLINERAA 359  
Db 322 GEATTINTGAKYADQLQIASLQINKMKSKAGCVTVGATLVADAKWSLTAEARLINERAA 381  
QY 360 HVSGQFRF 367  
Db 382 HVNAQFRF 389  
RESULT 7  
A40371  
major outer membrane protein precursor - Chlamydomophila psittaci (strain Fpn/pring)  
C:Species: Chlamydomophila psittaci, Chlamydia psittaci  
C:Date: 27-Nov-1991 #sequence\_revision 27-Nov-1991 #text\_change 31-Mar-2000  
C:Accession: A40371; S16137  
R:Storey, C.; Lusher, M.; Yates, P.; Richmond, S.  
J. Gen. Microbiol. 139, 2621-2626, 1993  
A:Title: Evidence for Chlamydia pneumoniae of non-human origin.  
A:Reference number: 140739; MUID:94103736  
A:Accession: 140859  
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL  
A:Residues: 1-392 <RES>  
A:Cross-references: EMBL:X61096; NID:940564; PIDN:CAA43409.1; PID:940565  
A:Experimental source: strain Fpn  
C:Genetics:  
A:Gene: MOMP  
C:Superfamily: Chlamydia major outer membrane protein  
C:Keywords: membrane protein  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-392/Product: major outer membrane protein #status predicted <MAT>  
Query Match 75.7%; Score 1477; DB 2; Length 392;  
Best Local Similarity 74.1%; Pred. No. 3e-112;  
Matches 275; Conservative 38; Mismatches 52; Indels 6; Gaps 4;  
QY 2 LPVGNPSPSLIDGTIWEAGAAGPCDPCATWCDALISIRAGFYGYVDFRILKVDPAKTF 61  
Db 23 LPVGNPAEPSSLIDGTIWEAGSGDPCDPCATWCDALISIRAGFYGYVDFRILKVDVKNKI 82  
QY 62 S-MGAKPTGSA--AANYTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRDVECTLGA 118  
Db 83 SCMAAPTAASGTANTVAADRSFAYGKHLDHAEFTNAGFIALNIWDRDVECTLGA 142  
QY 119 SNGYIRGNSTAFNLVGLFGVKGVTNNANLPVLSNGWVLYTDTFSWSVGARGALWE 178  
Db 143 SNGYKASSDAFNLVGLIGLACTDF -ANQRPVNEISQGLVELYDTTFSWSVGARGALWE 201  
QY 179 CCATILGAEFQYQAQSKPKVEELNVCNVSQSVNPKPKYKGA--FPLPTDAGVATATGTX 236  
Db 202 CCATILGAEFQYQAQSKPKVEELNVCNVSQSVNPKPKYKGA--FPLPTDAGVATATGTX 261  
QY 237 KSATINHEWQVGLSLVRLNLSVYICVQWNRATFDADNIRIAQPKLPTAVLNLTANNP 296  
Db 262 KSATINHEWQVGLSLVRLNLSVYICVQWNRATFDADNIRIAQPKLPTAVLNLTANNP 321  
QY 297 SLIGNATALSTTDSFDMQIVSCQINKFKSKAGCVTVGATLVADAKWSLTAEARLINE 356  
Db 322 TLLGVATTLDTNKNKYADFMQIVSMQINKMKSKAGCVTVGATLVADAKWSLTAEARLINE 381

QY 357 RAAHVSGQFRF 367  
Db 382 RAAHINAQFRF 392  
RESULT 8  
140740  
major outer membrane protein - Chlamydomophila psittaci (strain N352)  
C:Species: Chlamydomophila psittaci, Chlamydia psittaci  
C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 31-Mar-2000  
C:Accession: 140740  
R:Storey, C.; Lusher, M.; Yates, P.; Richmond, S.  
J. Gen. Microbiol. 139, 2621-2626, 1993  
A:Title: Evidence for Chlamydia pneumoniae of non-human origin.  
A:Reference number: 140739; MUID:94103736  
A:Accession: 140740  
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB  
A:Molecule type: DNA  
A:Residues: 1-402 <RES>  
A:Cross-references: GB:L04980; NID:9144544; PIDN:AAA17396.1; PID:9144545  
C:Genetics:  
A:Gene: momp  
C:Superfamily: Chlamydia major outer membrane protein  
Query Match 72.7%; Score 1419; DB 2; Length 402;  
Best Local Similarity 70.5%; Pred. No. 1.6e-107;  
Matches 268; Conservative 37; Mismatches 61; Indels 14; Gaps 6;  
QY 2 LPVGNPSPSLIDGTIWEAGAAGPCDPCATWCDALISIRAGFYGYVDFRILKVDPAKTF 61  
Db 23 LPVGNPAEPSSLIDGTIWEAGSGDPCDPCATWCDALISIRAGFYGYVDFRILKVDVKNKI 82  
QY 62 S-MGAKP---TCSAA--ANYTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRDVECTL 116  
Db 83 SCMAAPTAASGTANTVAADRSFAYGKHLDHAEFTNAGFIALNIWDRDVECTL 142  
QY 117 GASNGYIRGNSTAFNLVGLFGVKGVTNNANE---LPNVLSNGWVLYTDTFSWSVGA 172  
Db 143 GASNGYFKSSSAFNLVGLIGFSATSTSTELPMQLPNVGTQGVVEYDTFSWSVGA 202  
QY 173 RGALWECGATILGAEFQYQAQSKPKVEELNVCNVSQSVNPKPKYKGA--FPLPTDAGV 230  
Db 203 RGALWECGATILGAEFQYQAQSKPKVEELNVCNVSQSVNPKPKYKGA--FPLPTDAGV 262  
QY 231 ATATGTSKATINHEWQVGLSLVRLNLSVYICVQWNRATFDADNIRIAQPKLPTAVLN 290  
Db 263 TEATDTKSATINHEWQVGLSLVRLNLSVYICVQWNRATFDADNIRIAQPKLPTAVLN 322  
QY 291 LTANNPSLLGNATALSTT---DSFDMQIVSCQINKFKSKAGCVTVGATLVADAKWSL 347  
Db 323 ITTNWPSLLGSTALPNNAGKDVLSDLQIASIQINKMKSKAGCVTVGATLVADAKWSL 382  
QY 348 TAEARLINERAAHVSGQFRF 367  
Db 383 TGEARLINERAAHMAQFRF 402  
RESULT 9  
MNCWPM  
major outer membrane protein precursor - Chlamydomophila psittaci (strain A22/M)  
C:Species: Chlamydomophila psittaci, Chlamydia psittaci  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 31-Mar-2000  
C:Accession: S05954  
R:Pickett, M.A.; Everson, J.S.; Clarke, I.N.  
FEMS Microbiol. Lett. 55, 229-234, 1988  
A:Title: Chlamydia psittaci ewe abortion agent: complete nucleotide sequence of the  
A:Reference number: S05954  
A:Accession: S05954  
A:Molecule type: DNA  
A:Residues: 1-402 <PIC>  
A:Cross-references: EMBL:X12647; NID:940604; PIDN:CAA31177.1; PID:940605  
C:Superfamily: Chlamydia major outer membrane protein



F.1-22/Domain: signal sequence #status predicted <SIG>  
F.23-402/Product: major outer membrane protein #status predicted <MAT>

Query Match 72.5%; Score 1415; DB 1; Length 402;  
Best Local Similarity 70.3%; Pred. No. 3.4e-107;  
Matches 267; Conservative 37; Mismatches 62; Indels 14; Gaps 6;

QY 2 LPVGNPDPSSLIDGTIWEAGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDAPKTF 61  
DB 23 LPVGNPAEPLIDGTIWEAGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSA--ANTYTTAVDRNPAYNKHLDHAEFTNAGFTALNINWDFVFCFL 116  
DB 83 SGMAATPTQATGNASNTNPEANGRNIAIYGRHMDAEWFSNAAFALNINWDFDIFCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGYKGTIVNANE---LPNYSLSNGVVELYTDTSFSWSVGA 172  
DB 143 GASNGYFKSSAAFNVLVGLIGFSSTSTELPMQLPNVIGITQGVVEFTDTSFSWSVGA 202

QY 173 RGALWECGCATLGAEFQYQAQSKPKVEELNVICNVSOFSVKNPKGYKGA--FPLPTDAGV 230  
DB 203 RGALWECGCATLGAEFQYQAQSKPKIEVLNVTSSPAQFVIHKPRGYKGASSNFPPLITACT 262

QY 231 AFATGKTSATINHEWQVGSLSYRLNSLVPYIGVOWSRATEDADNIRIAQPKLPTAVLN 290  
DB 263 TEATDTKSAITIKYHEWQVGLSALYRLNMLVPYIGVWNSRATEDADTIRIAQPKLASEILN 322

QY 291 LTAWNPSSLGNATLSTT---DSFSDFMQIVSCQINKFKSRKACGVTVGCATLVADADKWSL 347  
DB 323 ITTNWPSLLGSTTLPNNNGKDVLSVLIQIASIQINKMSRKACGVAVGATLIDADKWSI 382

QY 348 TAEARLINERAHVSGQFRF 367  
DB 383 TGEARLINERAHMAHQFRF 402

RESULT 10  
B60109  
major outer membrane protein precursor - Chlamydia psittaci (strain meningopneumoni  
C:Species: Chlamydia psittaci, Chlamydia psittaci  
C:Date: 10-Nov-1992 #sequence\_revision 10-Nov-1992 #text\_change 31-Mar-2000  
C:Accession: B60109  
R:Zhang, Y.X.; Morrison, S.G.; Caldwell, H.D.; Baehr, W.  
Infect. Immun. 57, 1621-1625, 1989  
A:Title: Cloning and sequence analysis of the major outer membrane protein genes of two  
F.1-22/Domain: signal sequence #status predicted <SIG>  
F.23-389/Product: major outer membrane protein #status predicted <MAT>

A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-402 <ZHA>  
C:Superfamily: Chlamydia major outer membrane protein  
C:Keywords: membrane protein  
F.1-22/Domain: signal sequence #status predicted <SIG>  
F.23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 72.2%; Score 1410; DB 2; Length 402;  
Best Local Similarity 70.3%; Pred. No. 8.6e-107;  
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPDPSSLIDGTIWEAGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDAPKTF 61  
DB 23 LPVGNPAEPLIDGTIWEAGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSA--ANTYTTAVDRNPAYNKHLDHAEFTNAGFTALNINWDFVFCFL 116  
DB 83 SGMAATPTQATGNASNTNPEANGRNIAIYGRHMDAEWFSNAAFALNINWDFDIFCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGYKGTIVNANE---LPNYSLSNGVVELYTDTSFSWSVGA 172  
DB 143 GASNGYFKSSAAFNVLVGLIGFSSTSTELPMQLPNVIGITQGVVEFTDTSFSWSVGA 202

QY 173 RGALWECGCATLGAEFQYQAQSKPKVEELNVICNVSOFSVKNPKGYKGA--FPLPTDAGV 230  
DB 203 RGALWECGCATLGAEFQYQAQSKPKIEVLNVTSSPAQFVIHKPRGYKGASSNFPPLITACT 262

QY 231 AFATGKTSATINHEWQVGSLSYRLNSLVPYIGVOWSRATEDADNIRIAQPKLPTAVLN 290  
DB 263 TEATDTKSAITIKYHEWQVGLSALYRLNMLVPYIGVWNSRATEDADTIRIAQPKLASEILN 322

QY 291 LTAWNPSSLGNATLSTT---STDSFSDFMQIVSCQINKFKSRKACGVTVGCATLVADADKWSL 347  
DB 323 ITTNWPSLLGSTTLPNNNGKDVLSVLIQIASIQINKMSRKACGVAVGATLIDADKWSI 382

QY 348 TAEARLINERAHVSGQFRF 367  
DB 383 TGEARLINERAHMAHQFRF 402

RESULT 11  
A60341  
major outer membrane protein precursor - Chlamydia psittaci (strain 6BC)  
C:Species: Chlamydia psittaci, Chlamydia psittaci  
C:Date: 11-Dec-1992 #sequence\_revision 24-Feb-1994 #text\_change 31-Mar-2000  
C:Accession: A44565; A60341; B60341  
R:Everett, K.D.E.  
submitted to the EMBL Data Library, December 1990  
A:Reference number: A44565  
A:Accession: A44565  
A:Molecule type: DNA  
A:Residues: 1-402 <EVE>  
A:Cross-references: GB:X56980; NID:g40568; PIDN:CAA40300.1; PID:g40569  
R:Everett, K.D.E.; Andersen, A.A.; Plaunt, M.; Hatch, T.P.  
Infect. Immun. 59, 2853-2855, 1991  
A:Title: Cloning and sequence analysis of the major outer membrane protein gene of  
F.1-22/Domain: signal sequence #status predicted <SIG>  
F.23-402/Product: major outer membrane protein #status experimental <MAT>

Query Match 72.1%; Score 1408; DB 2; Length 402;  
Best Local Similarity 70.3%; Pred. No. 1.2e-106;  
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPDPSSLIDGTIWEAGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDAPKTF 61  
DB 23 LPVGNPAEPLIDGTIWEAGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSA--ANTYTTAVDRNPAYNKHLDHAEFTNAGFTALNINWDFVFCFL 116  
DB 83 SGMAATPTQATGNASNTNPEANGRNIAIYGRHMDAEWFSNAAFALNINWDFDIFCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGYKGTIVNANE---LPNYSLSNGVVELYTDTSFSWSVGA 172  
DB 143 GASNGYFKSSAAFNVLVGLIGFSSTSTELPMQLPNVIGITQGVVEFTDTSFSWSVGA 202

QY 173 RGALWECGCATLGAEFQYQAQSKPKVEELNVICNVSOFSVKNPKGYKGA--FPLPTDAGV 230  
DB 203 RGALWECGCATLGAEFQYQAQSKPKIEVLNVTSSPAQFVIHKPRGYKGASSNFPPLITACT 262

QY 231 AFATGKTSATINHEWQVGSLSYRLNSLVPYIGVOWSRATEDADNIRIAQPKLPTAVLN 290  
DB 263 TEATDTKSAITIKYHEWQVGLSALYRLNMLVPYIGVWNSRATEDADTIRIAQPKLASEILN 322

QY 291 LTAWNPSSLGNATLSTT---STDSFSDFMQIVSCQINKFKSRKACGVTVGCATLVADADKWSL 347



Db 323 ITTNPSLIGSTALPNNSGKDVLSVLQIASIQINKMSRKACGAVGATLIDADKWSI 382  
QY 348 TAEARLINERAAHVSGQFRF 367  
Db 383 TGEARLINERAAHVNAQFRF 402

## RESULT 12

C81747

major outer membrane protein, porin TC0052 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000

C:Accession: C81747

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255

A:Accession: C81747

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-387 <TEF>

A:Cross-references: GB:AE002272; GB:AE002160; NID:g7190083; PIDN:AAF38941.1; PID:g719009

A:Experimental source: strain Nigg (MoPn)

C:Genetics:

A:Gene: TC0052

C:Superfamily: Chlamydia major outer membrane protein

Query Match 69.2%; Score 1351.5; DB 2; Length 387;  
Best Local Similarity 66.4%; Pred. No. 4.5e-102;  
Matches 245; Conservative 50; Mismatches 67; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLLDGTIWEAGAGDPCDPCATWCDAISLRAGFYGYVDFRILKVDAPKTF 61  
Db 23 LPVGNPAEPLMIDGILWEGFGDPCDCTTWCDAISLRGLGYGVDFRVLKTDVNVKQF 82  
QY 62 SMGAKPTGSAANYTTA---VDPRNPAYNKHLDHAEWFTNAGFIALNIWDRFDFVCTLGA 118  
Db 83 EMGAFTGD--ADLTATPASRENPAKGMQDAEMFTNAYMALNIWDRFDFVCTLGA 140  
QY 119 SNGYIRGNSTAFNLVGLFGVKGTNNANLPVNSLNGVVELYDTFSWSVGARGALWE 178  
Db 141 TSGYLKGNASAFNLVGLFRDTEAADDIPNVLSQAQVVELYDTAFASVSGARAALWE 200  
QY 179 CCATILGAEFOYAQSKPKVEELNVCNVSQFVNPKPKYKGVAFPLPTDAGVATATGTS 238  
Db 201 CGCATLGASFOYAQSKPKVEELNVCNAAEFITNPKPKYGVQEFPLNPKAGTVSATDTKD 260  
QY 239 ATINHEWQVGSLSYRLNSLPYIGVQWVSRAFDADNIRIAQPKLPTAVLNLTAWNPSL 298  
Db 261 ASIDHEWQASLALSYRLNMFTPIYGVKWSRASFDADTIRIAQPKLPTAVLNLTAWNPTI 320  
QY 299 LGNATALSTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLINERA 358  
Db 321 SSGSIDVDLT--KITDTLQIVSLQNLKMSRKSCGLAIGTTIVADKYAVTVETRLIDERA 378  
QY 359 AHVSGQFRF 367  
Db 379 AHVNAQFRF 387

## RESULT 13

I40741

major outer membrane protein - Chlamydia trachomatis

C:Species: Chlamydia trachomatis

C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 20-Aug-1999

C:Accession: I40741

R:Zhang, Y.X.; Fox, J.G.; Ho, Y.; Zhang, L.; Stills, H.F.; Smith, T.F.

Mol. Biol. Evol. 10, 1327-1342, 1993

A:Title: Comparison of the major outer-membrane protein (MOMP) gene of mouse pneumonitis

A:Reference number: A49379; MUID:94104488

A:Accession: I40741  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-404 <RES>  
A:Cross-references: GB:L19221; NID:g410146; PIDN:AAAL6615.1; PID:g410147  
C:Genetics:  
A:Gene: momp  
C:Superfamily: Chlamydia major outer membrane protein

Query Match 69.2%; Score 1351; DB 2; Length 404;  
Best Local Similarity 65.0%; Pred. No. 5.3e-102;  
Matches 249; Conservative 50; Mismatches 66; Indels 18; Gaps 4;

QY 2 LPVGNPSPSLLDGTIWEAGAGDPCDPCATWCDAISLRAGFYGYVDFRILKVDAPKTF 61  
Db 23 LPVGNPAEPLMIDGILWEGFGDPCDCTTWCDAISLRGLGYGVDFRVLKTDVNVKQF 82  
QY 62 SMGAKPTGSG---AAANYTTAVDRPNPANKHLDHAEWFTNAGFIALNIWDRFDFVCTLGA 118  
Db 83 EMGPVPTTDDAADIITSTPRENPAYKGMQDAEMFTNAYMALNIWDRFDFVCTLGA 142  
QY 119 SNGYIRGNSTAFNLVGLFGVKG-----TTVNANLPVNSLNGVVELYDTFSWSV 170  
Db 143 TSGYLKGNASAFNLVGLFG-DGVANAANAIAATVAADSLPNVLSQAQVVELYDTAFANSV 201  
QY 171 GARGALWECGCATLGAEOYAQSKPKVEELNVCNVSQFVNPKPKYKGVAFPLPTDAGV 230  
Db 202 GARALWECGCATLGAFOYAQSKPKVEELNVCNAAEFTNPKPKYGVQEFPLALTAGT 261  
QY 231 ATATGKTGATINHEWQVGSLSYRLNSLPYIGVQWVSRAFDADNIRIAQPKLPTAVLN 290  
Db 262 DSATDTKASIDYHEWQASLALSYRLNMFTPIYGVKWSRASFDADTIRIAQPKLFAILD 321  
QY 291 LTANNPSLLGNAT-----ALSTDSFDMQIVSCQINKFKSRKACGVTVGATLVADAK 344  
Db 322 VTTWNPTTAGAGTTADGTGAATANGLATLQIVSLQNLKMSRKSCGLAIGTTIVADAK 381  
QY 345 WSLFAEARLINERAAHVSGQFRF 367  
Db 382 YAVTVETRLIDERAHVNAQFRF 404

## RESULT 14

JT0947

mouse pneumonitis major outer membrane protein - Chlamydia trachomatis

C:Species: Chlamydia trachomatis

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Aug-1999

C:Accession: JT0947

R:Fielder, T.J.; Pal, S.; Peterson, E.M.; de la Maza, L.M.

Gene 106, 137-138, 1991

A:Title: Sequence of the gene encoding the major outer membrane protein of the mouse

A:Reference number: JT0947; MUID:92039057

A:Accession: JT0947

A:Molecule type: DNA

A:Residues: 1-387 <FE>

A:Cross-references: GB:M64171; NID:g144536; PIDN:AAA23144.1; PID:g144537

C:Comment: C. trachomatis are gram negative bacteria that cause a variety of infections

C:Genetics:

A:Gene: om1MoPn

C:Superfamily: Chlamydia major outer membrane protein

C:Keywords: membrane protein

Query Match 69.1%; Score 1349.5; DB 2; Length 387;  
Best Local Similarity 66.1%; Pred. No. 6.6e-102;  
Matches 244; Conservative 50; Mismatches 68; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLLDGTIWEAGAGDPCDPCATWCDAISLRAGFYGYVDFRILKVDAPKTF 61  
Db 23 LPVGNPAEPLMIDGILWEGFGDPCDCTTWCDAISLRGLGYGVDFRVLKTDVNVKQF 82



QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEWFTNAGFIALNIWDREDFVCTLGA 118  
 DB 83 EMGAAPTGD--ADLTPTAPASRENPAYGKHQDAEMFTNAYMALNIWDREDFVCTLGA 140  
 QY 119 SNGYIRGNSTARNLGLFGVGTNNANLPNVLSNGVVELYTDTSFSWSVGARGALWE 178  
 DB 141 TSGYLKGNAAFNGLFGFGRDETAADIPNVLSQAQVVELYTDTAFAWSVGARGALWE 200  
 QY 179 CGCATLGAEFOYQAQSKPKVEELNVICNVSOFSVKNPKGYGVAFPLPTDAGVATATGKS 238  
 DB 201 CGCPTLGASFQQAQSKPKVEELNVLCNAEFTINPKGYGQEFPLNKGATVVSATDTKD 260  
 QY 239 ATINYHEWQVGSLSYRLNSLPYIGVQWRSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 298  
 DB 261 ASIDYHEWQASLSYRLNMFTPIYGVKWSRASFDADTIRIAQPKLETSILKMTTWNPTI 320  
 QY 299 LGNATALSTTDSFSDFMQIVSQINCKFKSRKACGVTGATLVADADKWSLTAEARLINERA 358  
 DB 321 SGSGIDVDT--KITDTLQIVSLQNLNKMKSRCGLAIGTTIVDADKYAVTVETRLIDERA 378  
 QY 359 AHVSGQFRF 367  
 DB 379 AHVNAQFRF 387

RESULT 15

SI6034  
 major outer membrane protein - Chlamydia trachomatis  
 C:Species: Chlamydia trachomatis  
 C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 12-Apr-1996  
 C:Accession: SI6034; A43875  
 R:Zhang, Y.X.; Zhang, L.  
 A:Submitted to the EMBL Data Library, July 1991  
 A:Description: The nucleotide sequence of major outer membrane protein gene of mouse bio  
 A:Reference number: SI6034  
 A:Accession: SI6034  
 A>Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-387 <ZHA>  
 A:Cross-references: EMBL:X60678  
 R:Peterson, E.M.; Cheng, X.; Markoff, B.A.; Fielder, T.J.; de la Maza, L.M.  
 Infect. Immun. 59, 4147-4153, 1991  
 A:Title: Functional and structural mapping of Chlamydia trachomatis species-specific ma  
 A:Reference number: A43875; MUID:92040090  
 A:Accession: A43875  
 A>Status: Preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 314-322 <PET>  
 A:Experimental source: mouse pneumonitis strain  
 A:Note: sequence extracted from NCBI backbone (NCBIP:62877)  
 C:Superfamily: Chlamydia major outer membrane protein

Query Match 69.1%; Score 1349.5; DB 2; Length 387;  
 Best Local Similarity 66.1%; Pred No. 6,6e-102;  
 Matches 244; Conservative 50; Mismatches 66; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLIDITWGAAGDPCDCAITWCDIAISLRAGFYGDYFDRILKVDAPKTF 61  
 DB 23 LPVGNPAEPLMIDITLWEGFGDPCDCTTWCDAISLRGLYGGDFVDRVLKTDVKNQF 82  
 QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEWFTNAGFIALNIWDREDFVCTLGA 118  
 DB 83 EMGAAPTGD--RDLTPTAPASRENPAYGKHQDAEMFTNAYMALNIWDREDFVCTLGA 140  
 QY 119 SNGYIRGNSTARNLGLFGVGTNNANLPNVLSNGVVELYTDTSFSWSVGARGALWE 178  
 DB 141 TSGYLKGNAAFNGLFGFGRDETAADIPNVLSQAQVVELYTDTAFAWSVGARGALWE 200  
 QY 179 CGCATLGAEFOYQAQSKPKVEELNVICNVSOFSVKNPKGYGVAFPLPTDAGVATATGKS 238  
 DB 201 CGCATLGASFQQAQSKPKVEELNVLCNAEFTINPKGYGQEFPLNKGATVVSATDTKD 260

QY 239 ATINYHEWQVGSLSYRLNSLPYIGVQWRSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 298  
 DB 261 ASIDYHEWQASLSYRLNMFTPIYGVKWSRASFDADTIRIGQPKLETSILKMTTWNPTI 320  
 QY 299 LGNATALSTTDSFSDFMQIVSQINCKFKSRKACGVTGATLVADADKWSLTAEARLINERA 358  
 DB 321 SGSGIDVDT--KITDTLQIVSLQNLNKMKSRCGLAIGTTIVDADKYAVTVETRLIDERA 378  
 QY 359 AHVSGQFRF 367  
 DB 379 AHVNAQFRF 387

Search completed: February 7, 2002, 21:38:11  
 Job time: 7100 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:42:47 ; Search time 76.51 seconds  
(without alignments)  
175.872 Million cell updates/sec

Title: US-09-391-606-16

Perfect score: 1952

Sequence: 1 MLFVGNPSDPSLLIDTWE.....TAEARLINERAAHVSGQFRF 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1947	99.7	389	1	OMPI_CHLPN P27455 chlamydia p
2	1882	96.4	389	1	OMIN_CHLPN Q07430 chlamydia p
3	1732	88.7	333	1	OMIK_CHLPN Q9xbfa chlamydia p
4	1534.5	78.6	389	1	OMIA_CHLPS P16567 chlamydia p
5	1477	75.7	392	1	OMIP_CHLPS Q00087 chlamydia p
6	1415	72.5	402	1	OMIE_CHLPS P10332 chlamydia p
7	1358.5	69.6	387	1	OMPI_CHLMU P75024 chlamydia m
8	1340.5	68.7	397	1	OMIH_CHLTR P13467 chlamydia t
9	1338.5	68.6	397	1	OMIN_CHLTR P23114 chlamydia t
10	1325	67.9	396	1	OMIA_CHLTR P23732 chlamydia t
11	1315.5	67.4	393	1	OMIE_CHLTR P17451 chlamydia t
12	1315.5	67.4	397	1	OMIC_CHLTR P08780 chlamydia t
13	1313.5	67.3	393	1	OMIL_CHLTR P19542 chlamydia t
14	1308.5	67.0	393	1	OMID_CHLTR Q46409 chlamydia t
15	1307.5	67.0	395	1	OMIF_CHLTR P16155 chlamydia t
16	1295	66.3	394	1	OMIM_CHLTR P06597 chlamydia t
17	1275	65.3	394	1	OMIB_CHLTR P23421 chlamydia t
18	1189	9.7	340	1	OMP2_CHLTR P38006 chlamydia t
19	182	9.3	340	1	OMP2_CHLMU Q9p113 chlamydia m
20	147.5	7.6	344	1	OMP2_CHLPN Q92752 chlamydia p
21	100.5	5.1	1919	1	HEX_ADE12 P19900 human adeno
22	98	5.0	1848	1	CBLA_CLOC P38058 clostridium
23	96	4.9	556	1	PHL_LEPIN P17627 leptospira
24	96	4.9	3421	1	TEGU_HSVB P28955 equine herp
25	95.5	4.9	864	1	AGLU_MUCJA Q92442 mucor javan
26	93.5	4.8	2021	1	OMPA_RICCN Q52657 rickettsia
27	93.5	4.8	2292	1	POLG_RICNV P17593 encephalomy
28	93.5	4.8	2292	1	POLG_EMCVD P17594 encephalomy
29	92.5	4.7	1169	1	YR82_YEAST P36170 saccharomyc
30	92.5	4.7	2290	1	POLG_EMCV P03304 encephalomy
31	91.5	4.7	639	1	AMYG_ASPPH P22832 aspergillus
32	91	4.7	448	1	FADL_ECOLI P10384 escherichia
33	90.5	4.6	1086	1	NNTM_MOUSE Q61941 mus musculu

## RESULT 1

ID	OMPI_CHLPN	STANDARD;	PRT;	389 AA.
AC	P27455; 09JOF6;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DE	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).			
GN	OMPA OR OMPI OR CPN0695 OR CP0051.			
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxID=83558;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=IOL-207;			
RX	MEDLINE=91237311; PubMed=2033374;			
RA	Carter M.W., Al-Mahdawi S.A.H., Giles I.G., Trehan J.D.,			
RA	Ward M.E., Clarke I.N.;			
RT	"Nucleotide sequence and taxonomic value of the major outer membrane			
RT	protein gene of Chlamydia pneumoniae IOL-207.";			
RL	J. Gen. Microbiol. 137:465-475(1991).			
RN	[2]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=TWAR;			
EX	MEDLINE=91244474; PubMed=1840574;			
RA	Perez Melgosa M., Kuo C.-C., Campbell L.A.;			
RT	"Sequence analysis of the major outer membrane protein gene of			
RT	Chlamydia pneumoniae.";			
RL	Infect. Immun. 59:2195-2199(1991).			
RN	[3]			
RC	SEQUENCE FROM N.A.			
RA	Mitchell W.M., Tharp A.C., Stratton C.W., Sriram S.;			
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=CNL029;			
EX	MEDLINE=99206606; PubMed=10192388;			
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,			
RA	Olinger L., Greenwood J., Davis R.W., Stephens R.S.;			
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";			
RL	Nat. Genet. 21:385-389(1999).			
RN	[5]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=AR39;			
EX	MEDLINE=20150255; PubMed=10684935;			
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,			
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,			
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,			
RA	Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,			
RA	Eisen J., Fraser C.M.;			
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia			
RT	pneumoniae AR39.";			
RL	Nucleic Acids Res. 28:1397-1406(2000).			
RN	[6]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=J138;			

34	90.5	4.6	1140	1	VHJL_ECOLI
35	90	4.6	533	1	FCY2_YEAST
36	89.5	4.6	726	1	FATA_VIBAN
37	89.5	4.6	867	1	SFMD_ECOLI
38	89.5	4.6	901	1	POLG_ENMG3
39	89	4.6	208	1	Y396_RICPR
40	89	4.6	363	1	YRAK_ECOLI
41	88.5	4.5	561	1	HLXB_PROMI
42	88.5	4.5	639	1	AMYG_ASPAK
43	87.5	4.5	482	1	PUR8_YEAST
44	87.5	4.5	1403	1	BIRF_MOUSE
45	87	4.5	664	1	PLBI_YEAST

## ALIGNMENTS

P37650	escherichia
P17064	saccharomyc
P11461	vibrio angu
P77468	escherichia
P32540	mengo encep
Q92dd5	rickettsia
P43319	escherichia
P16465	proteus mir
P23176	aspergillus
Q05911	saccharomyc
Q9Jib6	mus musculu
P39105	saccharomyc







```
QY 182 ATLGAEFOYAQSKPKVEELNVICNSQFSVKNPKGKGVAFPLPTDAGVATATGTSATI 241
|||||
Db 204 ATLGAEFOYAQSKPKVEELNVICNSQFSVKNPKGKGVAFPLPTDAGVATATGTSATI 263
|||||
QY 242 NYHEQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVLNLTANPSSLGN 301
|||||
Db 264 NYHEQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVLNLTANPSSLGN 323
|||||
QY 302 ATALSTTDSFSDPMQIVSQICQINKEKSKACGVTGATLVADKWSLTAEARLINRAAHV 361
|||||
Db 324 ATAVSSDOQFSDPMQIVSQICQINKEKSKACGVTGATLVADKWSLTAEARLINRAAHV 383
|||||
QY 362 SQQFRF 367
|||||
Db 384 SQQFRF 389
|||||

RESULT 3
OMIK_CHLPN STANDARD; PRT; 333 AA.
ID OM1K_CHLPN AC Q9XB4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN (MOMP) (FRAGMENT).
GN OMPA OR OMP1.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOALA TYPE I;
RX MEDLINE=93123168; PubMed=8419295;
RA Kaltenboeck B., Kousoules K.G., Storz J.;
RT "Structures of and allelic diversity and relationships among the major
RT outer membrane protein (OMP) genes of the four chlamydial species.";
RL J. Bacteriol. 175:487-502(1993).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M73038; AAD38210.1;
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin.
FT NON_TER 1
FT NON_TER 333
SQ SEQUENCE 333 AA; 35811 MW; 204604512C4C3B3F CRC64;
```

```
Query Match 88.7%; Score 1732; DB 1; Length 333;
Best Local Similarity 97.3%; Pred No. 3.2e-136;
Matches 324; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 23 AGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKFTSMGAKPTGSAANYTTAVDRP 82
|||||
Db 1 AGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKFTSMGAKPTGSAANYTTAVDRP 60
|||||

QY 83 NPAYNKLHDAEFNAGFIALNWDREDFVCTLGASNGYIRGNSTAFNLVGLFGVKGT 142
|||||
```

```
Db 61 NPAYNKLHDAEFNAGFIALNWDREDFVCTLGASNGYIRGNSTAFNLVGLFGVKGT 120
|||||
QY 143 VNANELPNVSLNSGVVELYTDTSFWSVGARGALWECCATLGAEFOYAQSKPKVEELNV 202
|||||
Db 121 VAANELPNVSLNSGVVELYTDTSFWSVGARGALWECCATLGAEFOYAQSKPKVEELNV 180
|||||
QY 203 ICNVQSFVNPKPGKGVAFPLPTDAGVATATGTSATINHEQVQASLSYRLNSLPVY 262
|||||
Db 181 ICNVAQFSVNPKPGKGVAFPLPTDAGVATATGTSATINHEQVQASLSYRLNSLPVY 240
|||||
QY 263 IGVQWSRATFDADNIRIAQPKLPTAVLNLTANPSSLGNATLSTDSFSDPMQIVSQICQ 322
|||||
Db 241 IGVQWSRATFDADNIRIAQPKLPTAVLNLTANPSSLGNATLSTDSFSDPMQIVSQICQ 300
|||||
QY 323 NKEKSKACGVTGATLVADKWSLTAEARLIN 355
|||||
Db 301 NKEKSKACGVTGATLVADKWSLTAEARLIN 333
|||||
```

```
RESULT 4
OM1A_CHLPS STANDARD; PRT; 389 AA.
ID OM1A_CHLPS AC P16567;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).
GN OMPA OR OMP1.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX MEDLINE=90128177; PubMed=2612883;
RA Herring A.J., Tan T.W., Baxter S., Inglis N.F., Dunbar S.;
RT "Sequence analysis of the major outer membrane protein gene of an
RT ovine abortion strain of Chlamydia psittaci.";
RL FEMS Microbiol. Lett. 53:153-158(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BOVINE ABORTION ISOLATE BAI;
RX MEDLINE=96189695; PubMed=8605581;
RA Griffiths P.C., Plater J.M., Martin T.C., Hughes S.L.,
RA Hughes K.J., Hewinson R.G., Dawson M.;
RT "Epizootic bovine abortion in a dairy herd: characterization of a
RT Chlamydia psittaci isolate and antibody response.";
RL Br. Vet. J. 151:683-693(1995).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X51859; CAA36152.1;
DR EMBL; L39020; AAB02850.1;
DR PIR; S08770; MMCWP3.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1
FT SIGNAL 22
```



FT CHAIN 23 389 MAJOR OUTER MEMBRANE PROTEIN.  
SQ SEQUENCE 389 AA: 41883 MW; 741B5A23ACDBB447 CRC64;

Query Match 78.6%; Score 1534.5; DB 1; Length 389;  
Best Local Similarity 75.7%; Pred. No. 8.8e-120;  
Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;

QY 2 LPVGNPSDPSLLIDGTIWEAGAAGDCPCATWCDALISIRAGYGYGVDFRILKVDAPKTF 61  
DB 23 LPVGNPAEPSSLIDGTIWEAGASGDCPCATWCDALISIRAGYGYGVDFRILKVDVKNKI 82

QY 62 S-MGAKPTGSA--AANYTTAVDRPNPAYNKLHDAEFTNAGFIALNIWDRFVCTLGASN 120  
DB 83 TGMGAVPTCTAAAYTKTPDRNPAYNKLHDAEFTNAGFIALNIWDRFVCTLGASN 142

QY 121 GYIRGNSTAFNLVGLFGVGTNNANLNPVLSNGVVELYDTDFSFWSVGARGALWECG 180  
DB 143 GYFKASSAFNLVGLFGVGTNNANLNPVLSNGVVELYDTDFSFWSVGARGALWECG 202

QY 181 CATLGAEFQYAOQKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSAT 240  
DB 203 CATLGAEFQYAOQKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSAT 262

QY 241 INYHEWQVGLSYRLNVLVPIGVQMSRATEDADNIRIAQPKLPTAVNLNLTAWRPSLLG 300  
DB 263 IYHEWQVGLSYRLNVLVPIGVQMSRATEDADNIRIAQPKLPTAVNLNLTAWRPSLLG 322

QY 301 NATALSTDTDFSDFMQIVSQINFKSKKACGVTGATLVADKWSLTAEARLINEAAH 360  
DB 323 EATLDTSNKFDADFLQIASIQINKMKSKKACGVTGATLVADKWSLTAEARLINEAAH 382

QY 361 VSGQFRF 367  
DB 383 MNAQFRF 389

RESULT 5  
OMIP\_CHLPS STANDARD; PRT; 392 AA.  
AC Q00087;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).  
GN OMPA OR OMP1.  
OS Chlamydia psittaci (Chlamydia psittaci).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.  
OX NCBI\_TaxID=83554;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94103736; PubMed=8277245;  
RA Storey C., Lusher M., Yates P., Richmond S.;  
RT "Evidence for Chlamydia pneumoniae of non-human origin.";  
RL J. Gen. Microbiol. 139:2621-2626(1993).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X61096; CAA43409.1; .

DR PIR: A40371; A40371.  
DR PIR: S16137; S16137.  
DR InterPro: IPR000604; Chlamydia\_OMP.  
DR Pfam: PF01308; Chlamydia\_OMP; 1.  
DR ProDom: PD001717; Chlamydia\_OMP; 1.  
KW Outer membrane; Transmembrane; Porin; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 392 MAJOR OUTER MEMBRANE PROTEIN.  
SQ SEQUENCE 392 AA: 42069 MW; 88B3C5D90BBA26DB CRC64;

Query Match 75.7%; Score 1477; DB 1; Length 392;  
Best Local Similarity 74.1%; Pred. No. 5.1e-115;  
Matches 275; Conservative 38; Mismatches 52; Indels 6; Gaps 4;

QY 2 LPVGNPSDPSLLIDGTIWEAGAAGDCPCATWCDALISIRAGYGYGVDFRILKVDAPKTF 61  
DB 23 LPVGNPAEPSSLIDGTIWEAGASGDCPCATWCDALISIRAGYGYGVDFRILKVDVKNKI 82

QY 62 S-MGAKPTGSA--AANYTTAVDRPNPAYNKLHDAEFTNAGFIALNIWDRFVCTLGASN 118  
DB 83 SGMAAPTAASCTASNTTVAADRSNFAKGLQDAEWCTNAAIYALNINWDRFVCTLGASN 142

QY 119 SNGYIRGNSTAFNLVGLFGVGTNNANLNPVLSNGVVELYDTDFSFWSVGARGALWEC 178  
DB 143 SNGYFKASSAFNLVGLFGVGTNNANLNPVLSNGVVELYDTDFSFWSVGARGALWEC 201

QY 179 CGCATLGAEFQYAOQKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSAT 236  
DB 202 CGCATLGAEFQYAOQKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSAT 261

QY 237 KSATINYNHEWQVGLSYRLNVLVPIGVQMSRATEDADNIRIAQPKLPTAVNLNLTAWRPS 296  
DB 262 KSATINYNHEWQVGLSYRLNVLVPIGVQMSRATEDADNIRIAQPKLPTAVNLNLTAWRPS 321

QY 297 SLLGNATLSTDTDFSDFMQIVSQINFKSKKACGVTGATLVADKWSLTAEARLINEAAH 356  
DB 322 TLLGVATTLDTSNKYADEFMQIVSQINFKSKKACGVTGATLVADKWSLTAEARLINEAAH 381

QY 357 RAAHVSGQFRF 367  
DB 382 RAAHINAQFRF 392

RESULT 6  
OMIP\_CHLPS STANDARD; PRT; 402 AA.  
AC P10332;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).  
GN OMPA OR OMP1.  
OS Chlamydia psittaci (Chlamydia psittaci).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.  
OX NCBI\_TaxID=83554;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=EAE A22/W;  
RA Pickett M.A., Everson S.J., Clarke I.N.;  
RT "Chlamydia psittaci ewe abortion agent: complete nucleotide sequence of the major outer membrane protein gene.";  
RL FEMS Microbiol. Lett. 55:229-234(1988).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X12647; CAA31177.1; -;  
DR EMBL; M36703; AAA23146.1; -;  
DR PIR; S05954; MNCWPM.  
DR InterPro: IPR000604; Chlamydia\_OMP.  
DR Pfam: PF01308; Chlamydia\_OMP; 1.  
DR ProDom: PD001717; Chlamydia\_OMP; 1.  
KW Outer membrane; Transmembrane; Porin; Signal;  
FT SIGNAL 1 22  
FT CHAIN 23 402 MAJOR OUTER MEMBRANE PROTEIN.  
SQ SEQUENCE 402 AA; 43277 MW; E6CF00D9DF1E87A CRC64;

Query Match 72.5%; Score 1415; DB 1; Length 402;  
Best Local Similarity 70.3%; Pred. No. 7.2e-110;  
Matches 267; Conservative 37; Mismatches 62; Indels 14; Gaps 6;

QY 2 LPVGNPSDPSLLIDGTWEGACGDCPCATWCDAISLRAGFYGVDFRILKVDAPKTF 61  
DB 23 LPVGNPAEPSLLIDGTWEGASGDCPCATWCDAISLRAGFYGVDFRILKVDVKNKF 82  
QY 62 S-MGAKP---TGSAA-ANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCFL 116  
DB 83 SGMAATPTQATGNASNTNQPENGARNIAYGRHQDAEWFSAFLALNIWDRFDFCFL 142  
QY 117 GASNGYIRGNSTAFNLVGLFGVKGTTVNANE-----LPNVSLNSGVVLYTDTFSFSVGA 172  
DB 143 GASNGYFKSSAAFNVLVGLIGESATSTSTELPMQPNVGVITQGVVEFTDTFSFSVGA 202  
QY 173 RGALWECGATGAEFYAQSKPKVEELVNICVSQFVSNKPKYKGA--PPLPDAGV 230  
DB 203 RGALWECGATGAEFYAQSKPKVEELVNICVSQFVSNKPKYKGA--PPLPDAGV 262  
QY 231 ATATGKTSATINHEWQVGSASLYRLNSLVPIYGVQMSRATFDADNIRIAQKPLTAVLN 290  
DB 263 TEATDTPKSAIKYHEWQVGLALSRLNMLVPIYGVQMSRATFDADNIRIAQKPLKSELN 322  
QY 291 LTAWNPISLGNATLSTT---DSFDFMQIVSQINKFKSRKACGVTVGATLVADKWSL 347  
DB 323 ITTWNPISLSTTTLPNNGKDLVSDVLAQIAQINKMKSRKACGVAVGATLVADKWSI 382  
QY 348 TAEARLINERAHVSGQFRF 367  
DB 383 TGEARLINERAHVSGQFRF 402

RESULT 7  
OMPL\_CHLMU STANDARD; PRT; 387 AA.  
AC P75024; Q04063; Q9X718;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).  
GN OMPA OR OMP1 OR TC0052.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MOPN:  
RX MEDLINE-92039057; PubMed-1937036;  
RA Fielder T.J., Pal S., Peterson E.M., la Maza L.M.;  
RT "Sequence of the gene encoding the major outer membrane protein of the  
mouse pneumonitis biovar of Chlamydia trachomatis.";  
RL Gene 106:137-138(1991).  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN-MOPN:  
RX MEDLINE-94104488; PubMed-8277858;  
RA Zhang Y.X., Fox J.G., Ho Y., Zhang L., Stills H.F., Smith T.F.;  
RT "Comparison of the major outer-membrane protein (MOMP) gene of mouse  
pneumonitis (MoPn) and hamster SFPP strains of Chlamydia trachomatis  
with other Chlamydia strains.";  
RL Mol. Biol. Evol. 10:1327-1342(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SSP.BV.MOUSE / NIGG II;  
RA Carter M.W., Giles I., Everson J.S., Clarke I.N.;  
RT "Chlamydia trachomatis mouse biovar: major outer membrane protein  
gene.";  
RL (In) Mardh P.A., la Placa M., Ward M. (eds.);  
Proceedings of the European society for chlamydia research and the  
second international symposium of Uppsala university centre for std  
research, pp.38-38, University of Uppsala, Uppsala (1992).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MOPN / NIGG;  
RX MEDLINE-20150255; PubMed-10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
White O., Hickey E.K., Peterson J., Utterback T., Berry K.,  
Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,  
Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G.,  
Salzberg S.L., Eisen J., Fraser C.M.;  
"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [5]  
RP SEQUENCE OF 37-375 FROM N.A.  
RC STRAIN-MOPN:  
RX MEDLINE-93123168; PubMed-8419295;  
RA Kaltenboeck B., Kousoulas K.G., Storz J.;  
RT "Structures of and allelic diversity and relationships among the major  
outer membrane protein (ompA) genes of the four chlamydial species.";  
RL J. Bacteriol. 175:487-502(1993).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: CELL WALL SURFACE.  
CC -!- MISCELLANEOUS: MOMP IS RESPONSIBLE FOR THE STRUCTURAL INTEGRITY OF  
THE EXTRA-CELLULAR INFECTIOUS ELEMENTARY BODY & THE DEVELOPMENTAL  
CONVERSION TO THE PLASTIC AND FRAGILE INTRACELLULAR RETICULATE  
BODY.  
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M64171; AAA23144.1; -;  
DR EMBL; U60196; AAB07068.1; -;  
DR EMBL; X63409; CAA45006.1; -;  
DR EMBL; AE002272; AAF38941.1; -;  
DR EMBL; M73044; AAD29101.1; -;  
DR TIGR; TC0052;  
DR InterPro: IPR000604; Chlamydia\_OMP.  
DR Pfam: PF01308; Chlamydia\_OMP; 1.  
DR ProDom: PD001717; Chlamydia\_OMP; 1.  
KW Outer membrane; Transmembrane; Porin; Signal; Complete proteome.  
FT SIGNAL 1 22 BY SIMILARITY.  
FT CHAIN 23 387 MAJOR OUTER MEMBRANE PROTEIN.  
FT CONFLICT 118 118 F -> Y (IN REF. 5).  
FT CONFLICT 123 123 Y -> F (IN REF. 5).  
FT CONFLICT 198 198 L -> P (IN REF. 1).  
FT CONFLICT 204 204 A -> P (IN REF. 1).



SQ SEQUENCE 387 AA; 42009 MW; 4FD6FDC23248E0A2 CRC64;

Query Match 69.6%; Score 1358.5; DB 1; Length 387;  
 Best Local Similarity 66.7%; Pred. No. 3.3e-105;  
 Matches 246; Conservative 50; Mismatches 66; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLLIDGTIWEAGAGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61  
 DB 23 LPVGNPAEPLMIDGILWEGFGDPCDCTTWCDAISLRLGYGDFVDFRILKVDVKNQF 82

QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEFTNAGFTALNIWDRFDFVCTLG 118  
 DB 83 EMGAFTGD--ADLTATPATSPRENPAAYGKHQDAEMFTNAYMALNIWDRFDFVCTLG 140

QY 119 SNGYIRGNSAFNLVGLFGVGTNNANELPNVSLNGSVWVLYDTFSWSVGARGALWE 178  
 DB 141 TSGYLKGNASAFNLVGLFGVGTNNANELPNVSLNGSVWVLYDTFSWSVGARGALWE 200

QY 179 CGCATLGAFFOYAKSKPKVEELNVCNVSQFVNPKYKGVAPPLPTDAGVATATGKS 238  
 DB 201 CCATLGAFFOYAKSKPKVEELNVCNVSQFVNPKYKGVAPPLPTDAGVATATGKS 260

QY 239 ATINHEWQVGLSLYRLNLSVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAWNPSL 298  
 DB 261 ASIDYHEWQVGLSLYRLNLSVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAWNPSL 320

QY 299 LGNATALSTDSFSDPMQIVSQINPKFKSKAGCVTVGATLVADKWSLTAEARLINERA 358  
 DB 321 SSGSDIVDT--KITDTLQIVSLQNLNMRKSRKSCGLAIGTTIVDADKYAVTVETRIDERA 378

QY 359 AHVSGQFRF 367  
 DB 379 AHVNAQFRF 387

RESULT 8  
 OM1H.CHLTR STANDARD; PRT; 397 AA.

AC P13467;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR H PRECURSOR (MOMP).  
 GN OMPA OR OMP1H.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=SEROVAR H;  
 RX MEDLINE=90045958; PubMed=2813066;  
 RA Hamilton P.T., Malinowski D.P.;  
 RT "Nucleotide sequence of the major outer membrane protein gene from  
 Chlamydia trachomatis serovar H.";  
 RL Nucleic Acids Res. 17:8366-8366(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H/UW-4;  
 RA Dean D.A.;  
 RT "Sequence analysis of the major outer membrane protein gene (ompA) of  
 Chlamydia trachomatis.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
 BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
 THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
 MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
 CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X16007; CAA34145.1;  
 EMBL; AF304857; AAG41415.1;  
 PIR; S06589; MNCWTH.  
 InterPro; IPR000604; Chlamydia\_OMP.  
 Pfam; PF01308; Chlamydia\_OMP; 1.  
 ProDom; PD001717; Chlamydia\_OMP; 1.  
 Outer membrane; Transmembrane; Porin; Signal.  
 SIGNAL 1 22  
 CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR H.  
 FT CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR H.  
 SQ SEQUENCE 397 AA; 42946 MW; 478ACE3808BF37BA CRC64;

Query Match 68.7%; Score 1340.5; DB 1; Length 397;  
 Best Local Similarity 65.1%; Pred. No. 1e-103;  
 Matches 244; Conservative 51; Mismatches 71; Indels 9; Gaps 3;

QY 2 LPVGNPSPSLLIDGTIWEAGAGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61  
 DB 23 LPVGNPAEPLMIDGILWEGFGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDVKNKEF 82

QY 62 SMGAKPTGSAANY---TTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFDFVCTLG 117  
 DB 83 QMGAFTTNDADLNDPKTNVARNPAYGKHQDAEMFTNAYMALNIWDRFDFVCTLG 142

QY 118 ASGYIRGNSAFNLVGLFGVGTNNANELPNVSLNGSVWVLYDTFSWSVGARG 174  
 DB 143 ATTYGLKGNASAFNLVGLFGVGTNNANELPNVSLNGSVWVLYDTFSWSVGARG 202

QY 175 ALWECGCATLGAFFOYAKSKPKVEELNVCNVSQFVNPKYKGVAPPLPTDAGVATAT 234  
 DB 203 ALWECGCATLGAFFOYAKSKPKVEELNVCNVSQFVNPKYKGVAPPLPTDAGVATAT 262

QY 235 GTSKATINHEWQVGLSLYRLNLSVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 294  
 DB 263 GTSKATINHEWQVGLSLYRLNLSVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 322

QY 295 NPSLGNAT--ALSTDSFSDPMQIVSQINPKFKSKAGCVTVGATLVADKWSLTAEAR 352  
 DB 323 NPTIAGKTGVASGSDNLDADTMQIVSLQNLNMRKSRKSCGLAIGTTIVDADKYAVTVETR 382

QY 353 LINERAAHVSGQFRF 367  
 DB 383 LIDERAAHVNAQFRF 397

RESULT 9  
 OM1H.CHLTR STANDARD; PRT; 397 AA.

AC P23114;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3 PRECURSOR (MOMP).  
 GN OMPA OR OMP1L3.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=404 / SEROVAR L3;  
 RX MEDLINE=91285429; PubMed=2060793;  
 RA Fieldner T.J., Peterson E.M., de la Maza L.M.;  
 RT "Nucleotide sequence of DNA encoding the major outer membrane protein  
 of Chlamydia trachomatis serovar L3.";  
 RL Gene 101:159-160(1991).  
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
 BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH



CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
 CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X55700; CAA39226.1; -  
 CC PIR: JE0413; JE0413.  
 CC InterPro: IPR000604; Chlamydia\_OMP.  
 CC Pfam: PF01308; Chlamydia\_OMP; 1.  
 CC ProDom: PD001717; Chlamydia\_OMP; 1.  
 CC Outer membrane; Transmembrane; Porin; Signal.  
 CC SIGNAL 1 22 BY SIMILARITY.  
 CC CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3.  
 CC SEQUENCE 397 AA; 42885 MW; FIDDCF09535C2595 CRC64;

Query Match 68.6%; Score 1338.5; DB 1; Length 397;  
 Best Local Similarity 65.1%; Pred. No. 1.5e-103;  
 Matches 244; Conservative 49; Mismatches 73; Indels 9; Gaps 3;

QY 2 LPVGNPSPSLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGYDFVDRILKVDAPKTF 61  
 DB 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISMRVGYGDFVDRVLKTDVNKEF 82  
 QY 62 SNGAKPTGSAANY-----TTAVDRNPAYNKLHDAEFTNAGFTALNIWDRFVFCITLG 117  
 DB 83 QMGAEPTTSDAGLSNDPTTNVARNPAYGKHQDAEFTNAAAYMALNIWDRFVFCITLG 142  
 QY 118 ASNGIRGNSAFNLVGLFGVKGTTVNANE---LPNVSLNSGVVELYDTTFSWSVGARG 174  
 DB 143 ATTGVLKGNASAFNLVGLFGTKTQSTNFNTAKLVPTALNOAVVELYDTTTFANVSVGARA 202  
 QY 175 ALWEGCGATLGAFOYAGSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATAT 234  
 DB 203 ALWEGCGATLGAFOYAGSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATAT 262  
 QY 235 GTSKATINHEWQVGSLSYRLNLPVYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 294  
 DB 263 GTKDASIDYHEWQASLSYRLNLPVYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 322  
 QY 295 NPSLL--GNATLSTDSFSDFMQIVSCQINPKFKSRKAGCVTVGATLVADKWSLTAEARL 352  
 DB 323 NPTIAGKGSVVASGSENELADTMQIVSLQLNKKMSRKSGIAGVTIYDADKYAVTVETR 382  
 QY 353 LINERAAHVSGQFRF 367  
 DB 383 LIDERAHVNAQFRF 397

RESULT 10  
 OMIA\_CHLTR STANDARD; PRT; 396 AA.  
 ID OMIA\_CHLTR  
 AC P23732;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR A PRECURSOR (MOMP).  
 GN OMPA OR OMP1A.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SAI/OT / SEROVAR A;

RX MEDLINE=91045088; PubMed=2235504;  
 RA Hayes L.J., Clarke I.N.;  
 RT "Nucleotide sequence of the major outer membrane protein gene of  
 RT Chlamydia trachomatis strain A/Sai/OT";  
 RL Nucleic Acids Res. 18:6136-6136(1990).  
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
 CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
 CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
 CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M58938; AAA23141.1; -  
 CC EMBL: N33635; AAA92785.1; -  
 CC PIR: S12799; S12799.  
 CC InterPro: IPR000604; Chlamydia\_OMP.  
 CC Pfam: PF01308; Chlamydia\_OMP; 1.  
 CC ProDom: PD001717; Chlamydia\_OMP; 1.  
 CC Outer membrane; Transmembrane; Porin; Signal.  
 CC SIGNAL 1 22  
 CC CHAIN 23 396 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR A.  
 CC SEQUENCE 396 AA; 42877 MW; 2F9D3H0CE2D08162 CRC64;

Query Match 67.9%; Score 1325; DB 1; Length 396;  
 Best Local Similarity 64.7%; Pred. No. 2e-102;  
 Matches 242; Conservative 47; Mismatches 77; Indels 8; Gaps 3;

QY 2 LPVGNPSPSLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGYDFVDRILKVDAPKTF 61  
 DB 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISMRVGYGDFVDRVLKTDVNKEF 82  
 QY 62 SNGAKPTGSAANY-----TTAVDRNPAYNKLHDAEFTNAGFTALNIWDRFVFCITLG 117  
 DB 83 QMGAEPTTSDAGLEKDPVNVARNPAYGKHQDAEFTNAAAYMALNIWDRFVFCITLG 142  
 QY 118 ASNGIRGNSAFNLVGLFGVKGTTV---NANELPNVSLNSGVVELYDTTFSWSVGARG 174  
 DB 143 ATTGVLKGNASAFNLVGLFGTKTQSGGDTANIVENTALNOAVVELYDTTTFANVSVGARA 202  
 QY 175 ALWEGCGATLGAFOYAGSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATAT 234  
 DB 203 ALWEGCGATLGAFOYAGSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATAT 262  
 QY 235 GTSKATINHEWQVGSLSYRLNLPVYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 294  
 DB 263 GTKDASIDYHEWQASLSYRLNLPVYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 322  
 QY 295 NPSLLGNATLSTDS--FSDFMQIVSCQINPKFKSRKAGCVTVGATLVADKWSLTAEARL 353  
 DB 323 NPTIAGKGSVVASGSENELADTMQIVSLQLNKKMSRKSGIAGVTIYDADKYAVTVETR 382  
 QY 354 LINERAAHVSGQFRF 367  
 DB 383 LIDERAHVNAQFRF 396

RESULT 11  
 OMIA\_CHLTR STANDARD; PRT; 393 AA.  
 ID OMIA\_CHLTR  
 AC P17451;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)



DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR E PRECURSOR (MOMP).  
GN OMPA OR OMPLE.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BOUR / SEROVAR E;  
RX MEDLINE=90287737; PubMed=2356137;  
RA Peterson E.M., Markoff B.A., de la Maza L.M.;  
RT "The major outer membrane protein nucleotide sequence of Chlamydia trachomatis, serovar E.";  
RL Nucleic Acids Res. 18:3414-3414(1990).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTIONS THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X52557; CAA36791.1; -  
DR PIR; S10201; MWCWTE.  
DR InterPro: IPR000604; Chlamydia\_OMP.  
DR Pfam; PF01308; Chlamydia\_OMP; 1.  
DR ProDom; PD001717; Chlamydia\_OMP; 1.  
KW Outer membrane; Transmembrane; Porin; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR E.  
SQ SEQUENCE 393 AA; 42424 MW; AB2B82D16027B361 CRC64;  
  
Query Match 67.4%; Score 1315.5; DB 1; Length 393;  
Best Local Similarity 65.3%; Pred. NO. 1.2e-101;  
Matches 243; Conservative 46; Mismatches 76; Indels 7; Gaps 4;  
  
QY 2 LPVGNPSPSLIDGTINWEGAGDCPCATWCDALISLRAGFYGDYVDFDRILKVDPAKTF 61  
DB 23 LPVGNPAEPLSMIDGILWEGFGDCPCCTWCDALISLRAGFYGDYVDFDRILKVDYNEF 82  
  
QY 62 SMGAKP---TGSAAYNTTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRDFVCTGLA 118  
DB 83 QMGDKPTSTTGNATAP-TTLTARENPAYGRHMQDAEMFTNAACMALNIWDRDFVCTGLA 141  
  
QY 119 SNGYIRGNSTAFNLVGLFG--VKGTIVANLPLPNVSLNGVVELYTDTSFSSVSGARGAL 176  
DB 142 SSGYLKGNASFNVLGFGDNEQSTVNTSPNNSLDQSVVELYTDTSFSSVSGARGAL 201  
  
QY 177 WCGCATLCAEFOYQAQSKPKVEELNVCNVSFNKPKYKGVAFPLPTDAGVATATGT 236  
DB 202 WCGCATLCAEFOYQAQSKPKVEELNVCNVAEFTINKPKYGVQEPFLAINGTDAATGT 261  
  
QY 237 KSATINHEWQVQASLSYRLNSLVPYIGVQWVSRAFDADNIRIAQKPLTAVLNITANNP 296  
DB 262 KDAIDYHEWQASLASYRLNMFPTPIGVKWSRASFDADTIRIAQPKSAIFDITLNP 321  
  
QY 297 SLLGNATLSTTD-SFSDPMQIVSQINKFKSKAGGVTVGATLVADKWSITAEARLIN 355  
DB 322 TTAGAGDVKASAEQGLGDMQIVSLQNLNKKSRKSCGTAVGTTIVDADKAVTVETRLID 381  
  
QY 356 ERAAHVSGQFRF 367  
DB 382 ERAAHVNAQFRF 393

RESULT 12  
OMC\_CHLTR ID OMC\_CHLTR STANDARD; PRT; 397 AA.  
AC P08780;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR C PRECURSOR (MOMP).  
GN OMPA OR OMP1C.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=87307955; PubMed=3040664;  
RX Stephens R.S., Sanchez-Pescador R., Wagar E.A., Inouye C., Urdea M.S.;  
RA "Diversity of Chlamydia trachomatis major outer membrane protein genes.";  
RT J. Bacteriol. 169:3879-3885(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C/TW3;  
RX MEDLINE=20407420; PubMed=10950788;  
RA Dean D., Suchland R.J., Stamm W.E.;  
RT "Evidence for long-term cervical persistence of Chlamydia trachomatis by omp1 genotyping.";  
RL J. Infect. Dis. 182:909-916(2000).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTIONS THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M17343; AAA23156.1; -  
DR EMBL; AF202455; AAG09443.1; -  
DR PIR; S11011; MWCWTC.  
DR InterPro: IPR000604; Chlamydia\_OMP.  
DR Pfam; PF01308; Chlamydia\_OMP; 1.  
DR ProDom; PD001717; Chlamydia\_OMP; 1.  
KW Outer membrane; Transmembrane; Porin; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR C.  
SQ SEQUENCE 397 AA; 42892 MW; 0047BCDB108E5309 CRC64;  
  
Query Match 67.4%; Score 1315.5; DB 1; Length 397;  
Best Local Similarity 64.3%; Pred. NO. 1.2e-101;  
Matches 241; Conservative 49; Mismatches 76; Indels 9; Gaps 3;  
  
QY 2 LPVGNPSPSLIDGTINWEGAGDCPCATWCDALISLRAGFYGDYVDFDRILKVDPAKTF 61  
DB 23 LPVGNPAEPLSMIDGILWEGFGDCPCCTWCDALISLRAGFYGDYVDFDRILKVDYNEF 82  
  
QY 62 SMGAKPTGSAANY----TTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRDFVCTGL 117  
DB 83 QMGAAPTSDVAGLQNDPTINVARPNPAYKHMQDAEMFTNAAVWALNIWDRDFVCTGL 142  
  
QY 118 ASNGYIRGNSTAFNLVGLFGVKGTTVNANE----LPNVSLNGVVELYTDTSFSSVSGARG 174  
DB 143 ATTGYLKNASFNVLVGLFGTQSSSFNTAKLIPNTALNEAVVELYINTTFAWSVGARA 202  
  
QY 175 ALWEGCATLGAEFOYQAQSKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATAT 234



Db 203 ALWECGCATLGASFOYAKSPKVEELNVLNASEFTINKPKYGVGAEPFLNITAGTEAAT 262  
Qy 235 GTSKATINHEWQVGSLSYRLNVLVPIGVQWSRATDADNIRIAQPKLPTAVLNLTAW 294  
Db 263 GTSKASIDYHEWQASLSYRLNVLVPIGVQWSRATDADNIRIAQPKLAEIUDVTTL 322  
Qy 295 NPSLGNATALS--TTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEAR 352  
Db 323 NRTTAGKSVVSAGTVDNELADTMQIVSLQNLNKMKSCKGIAVGTITVDADKYAVTVEAR 382  
Qy 353 LINERAAHVSGQFR 367  
Db 383 LIDERRAAHVNAQFR 397

RESULT 13  
OM1D\_CHLTR STANDARD; PRT; 393 AA.  
AC P19542;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L1 PRECURSOR (MOMP).  
GN OMPA OR OMP1L.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Pickett M.A., Ward M.E., Clarke I.N.;  
RT "Complete nucleotide sequence of the major outer membrane protein  
gene from Chlamydia trachomatis serovar L1.";  
RL FEMS Microbiol. Lett. 42:185-190(1987).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M36533; AAA23142.1; -  
DR PIR; S06259; S06259.  
DR InterPro; IPR000604; Chlamydia\_OMP.  
DR Pfam; PF01308; Chlamydia\_OMP; 1.  
DR ProDom; PD00117; Chlamydia\_OMP; 1.  
KW Outer membrane; Transmembrane; Porin; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L1.  
SQ SEQUENCE 393 AA; 42543 MW; 7A952839408EE2DF CRC64;

Query Match 67.3%; Score 1313.5; DB 1; Length 393;  
Best Local Similarity 64.5%; Pred. No. 1.8e-101;  
Matches 240; Conservative 51; Mismatches 74; Indels 7; Gaps 4;

Qy 2 LPVGNPSPSLIDGTIWEGAAGDPCDPCATWCDAISLRAGFYGVDFRILKVDAPKTF 61  
Db 23 LPVGNPAEPRLMIDGILWEGFGDPCDPCDPCATWCDAISLRAGFYGVDFRILKVDAPKTF 82

Qy 62 SNGAKP---TCSAANYTTAVDRPNPAYNKLHDAEWNTNAGFTALNTWDRFDVCTLGA 118  
Db 83 QMGAKPTATTGNAAP--STCTARENPAYGRHMQDAEMFTNAYMALNTWDRFDVCTLGA 141

Qy 119 SNGYIRGRNSTAFNLVGLFG--VKGTTVNANLPNVSLNSGVVELYTDTSFMSVSGARGAL 176

Db 142 TSGYLKNGSASFNLVGLFGDNEQSTYKKNDAVPNMSFDQSVVELYTDITTFMSVGAARAL 201  
Qy 177 WECCGATLGAFQYAKSPKVEELNVLNVCNVSQFSVKNPKYGVAFPLPTDAGVATATGT 236  
Db 202 WECCGATLGAFQYAKSPKVEELNVLNVCNVCNVSQFSVKNPKYGVAFPLPTDAGVATATGT 261  
Qy 237 KSATINHEWQVGSLSYRLNVLVPIGVQWSRATDADNIRIAQPKLPTAVLNLTAW 296  
Db 262 KDSIDYHEWQASLSYRLNVLVPIGVQWSRATDADNIRIAQPKLPTAVLNLTAW 321  
Qy 297 SLLGNATALTSD--SFSDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLN 355  
Db 323 TIAGAGEVKAEGQLGDMQIVSLQNLNKMKSCKGIAVGTITVDADKYAVTVETRLID 381  
Qy 356 EFAAHVSGQFR 367  
Db 382 EFAAHVNAQFR 393

RESULT 14  
OM1D\_CHLTR STANDARD; PRT; 393 AA.  
AC Q46409;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR D PRECURSOR (MOMP).  
GN OMPA OR OMP1 OR CT681.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=D/B-120;  
RC MEDLINE=93013014; PubMed=1398119;  
RA Sayada C., Denamur E., Elion J.;  
RT "Complete sequence of the major outer membrane protein-encoding gene  
of Chlamydia trachomatis serovar Da.";  
RL Gene 120:129-130(1992).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC STRAIN=D/IU-71960;  
CC MEDLINE=98339860; PubMed=9673241;  
CC Stothard D.R., Boguslawski G., Jones R.B.;  
CC "Phylogenetic analysis of the Chlamydia trachomatis major outer  
membrane protein and examination of potential pathogenic  
determinants.";  
CC Infect. Immun. 66:3618-3625(1998).  
CC [3]  
CC SEQUENCE FROM N.A.  
CC STRAIN=D/UW-3/CX;  
CC MEDLINE=99000809; PubMed=9784136;  
CC Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
CC Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
CC Davis R.W.;  
CC "Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis.";  
CC Science 282:754-759(1998).  
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----



or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL; X62918; CAA44701.1; -  
 CC EMBL; AF063195; AC31436.2; -  
 CC EMBL; AE001338; AC68276.1; -  
 CC InterPro; IPR000604; Chlamydia\_OMP.  
 CC Pfam; PF01308; Chlamydia\_OMP; 1.  
 CC ProDom; PD001717; Chlamydia\_OMP; 1.  
 CC Outer membrane; Transmembrane; Porin; Signal; Complete proteome.  
 KW SIGNAL 1 22 BY SIMILARITY  
 FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR D.  
 SQ SEQUENCE 393 AA; 42438 MW; 8CD692FD3EFF21D6 CRC64;

Query Match 67.0%; Score 1308.5; DB 1; Length 393;  
 Best Local Similarity 64.8%; Pred. No. 4.6e-101;  
 Matches 241; Conservative 48; Mismatches 76; Indels 7; Gaps 4;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 DB 23 LPVGNPAEPLMIDGILWEGFGDPCDPCATWCDAISMRVGYGDFVDRVLKTDVNFKEF 82  
 QY 62 SWGAKP---TGAANYTTAVDRNPAYNKHLHDAEFTNAGFTALNIWDRFDFVCTLGA 118  
 DB 83 QMGAKPTTDTGNSAAP-STLTARENPAYGRHMQDAEMFTNAACMALNIWDRFDFVCTLGA 141  
 QY 119 SNGYIRGNSTAFNLVGLFG--VKGTNNANLNPVSLNSGVVELYTDTSFSSWSVGARGAL 176  
 DB 142 TSGYLKGNASFNVLGLFGDGNENQKTVKASVPNNSFDSQSVVELYTDITTFAWSVGARGAL 201  
 QY 177 WECGCATLGAFQYQAQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGT 236  
 DB 202 WECGCATLGASFOYQAQSKPKVEELNVCNVAEFTINKPKGYGVKEFFPLDLTAGTDAATGT 261  
 QY 237 KSATINHEWQVGSLSYRLNSLPYIGVQWSRATFDADNIRIAOPKPLTAVLNLTAWN 296  
 DB 262 KDASIDYHEWQASLSYRLNMFPTYGKWSRASFDADTIRIAQPSATAIFDTITLNP 321  
 QY 297 SLGNATALSTDD--SFSDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLN 355  
 DB 322 TIAGAGDYKTAGEGQDGTMOIVSLQLNKMRSKSGIAVGTITVDADKYAVTVETRLID 381  
 QY 356 ERAAHVSGQFRF 367  
 DB 382 ERAAHVNAQFRF 393

RESULT 15  
 OMIF\_CHLTR STANDARD; PRT; 395 AA.  
 ID OMIF\_CHLTR  
 AC P16155;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR F PRECURSOR (MOMP).  
 GN OMPA OR OMPF.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IC-CAL3 / SEROVAR F;  
 RX MEDLINE=90192102; PubMed=2315025;  
 RA Zhang Y.X., Morrison S.G., Caldwell H.D.;  
 RT "The nucleotide sequence of major outer membrane protein gene of  
 Chlamydia trachomatis serovar F";  
 RL Nucleic Acids Res. 18:1061-1061(1990).  
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
 CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
 CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.

CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

DR EMBL; X52080; CAA36299.1; -  
 DR InterPro; IPR000604; Chlamydia\_OMP.  
 DR Pfam; PF01308; Chlamydia\_OMP; 1.  
 DR ProDom; PD001717; Chlamydia\_OMP; 1.  
 KW Outer membrane; Transmembrane; Porin; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 395 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR F.  
 SQ SEQUENCE 395 AA; 42586 MW; 7F90FFDEEC264ACF CRC64;

Query Match 67.0%; Score 1307.5; DB 1; Length 395;  
 Best Local Similarity 63.5%; Pred. No. 5.6e-101;  
 Matches 237; Conservative 53; Mismatches 76; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 DB 23 LPVGNPAEPLMIDGILWEGFGDPCDPCATWCDAISMRVGYGDFVDRVLKTDVNFKEF 82  
 QY 62 SMG---AKPTGSAANYTTAVDRNPAYNKHLHDAEFTNAGFTALNIWDRFDFVCTLGA 118  
 DB 83 EMGEALAGASGNTTSTLSKLVERTNPAYGKHMQDAEMFTNAACMTLNIWDRFDFVCTLGA 142  
 QY 119 SNGYIRGNSTAFNLVGLFG--GVKGTNNANLNPVSLNSGVVELYTDTSFSSWSVGARGAL 176  
 DB 143 TSGYLKGNASFNVLGLFGDGNVNAKPAADSIQNVQLNSVVELYTDITTFAWSVGARGAL 202  
 QY 177 WECGCATLGAFQYQAQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGT 236  
 DB 203 WECGCATLGASFOYQAQSKPKVEELNVCNVAEFTINKPKGYGVKEFFPLDLTAGTDAATGT 262  
 QY 237 KSATINHEWQVGSLSYRLNSLPYIGVQWSRATFDADNIRIAOPKPLTAVLNLTAWN 296  
 DB 263 KDASIDYHEWQASLSYRLNMFPTYGKWSRASFDADTIRIAQPSATAIFDTITLNP 322  
 QY 297 SL--LGNATALSTDD--SFSDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLI 354  
 DB 323 TIAGCGSVAGANTEGQISDTHQIVSLQLNKMRSKSGIAVGTITVDADKYAVTVETRLI 382  
 QY 355 NERAAHVSGQFRF 367  
 DB 383 DERAHVNAQFRF 395

Search completed: February 7, 2002, 21:42:47  
 Job time: 601 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:41:21 ; Search time 172 Seconds  
(without alignments)  
312.104 Million cell updates/sec

Title: US-09-391-606-16

Perfect score: 1952

Sequence: 1 MLPVGNPSPDLLDGTWE.....TAEARLINERAHVSGQFRF 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL\_17.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1910	97.8	389	2 Q08085	Q08085 chlamydia p
2	1531.5	78.5	389	2 Q9APM4	Q9APM4 chlamydophi
3	1529.5	78.4	388	2 Q9AIK1	Q9AIK1 chlamydia p
4	1501.5	76.9	391	2 Q46235	Q46235 chlamydia p
5	1497.5	76.7	388	2 Q9AIK0	Q9AIK0 chlamydia p
6	1497.5	76.7	389	2 Q9AIH9	Q9AIH9 chlamydophi
7	1487	76.2	392	2 Q9AIJ4	Q9AIJ4 chlamydia p
8	1480	75.8	381	2 Q9AIJ2	Q9AIJ2 chlamydophi
9	1479	75.8	390	2 Q9AIJ5	Q9AIJ5 chlamydia p
10	1477	75.7	392	2 Q99QB0	Q99QB0 chlamydophi
11	1469.5	75.3	380	2 Q9AIJ1	Q9AIJ1 chlamydophi
12	1423	72.9	391	2 Q9AIJ2	Q9AIJ2 chlamydia p
13	1420.5	72.8	341	2 Q9X717	Q9X717 chlamydophi
14	1419	72.7	402	2 Q46193	Q46193 chlamydia p
15	1415	72.5	382	2 Q9AIJ9	Q9AIJ9 chlamydia p
16	1415	72.5	395	2 Q9AIJ7	Q9AIJ7 chlamydia p
17	1415	72.5	397	2 Q9AIJ8	Q9AIJ8 chlamydia p
18	1413	72.4	402	2 Q9AIJ6	Q9AIJ6 chlamydia p
19	1410	72.2	402	2 Q9AIJ0	Q9AIJ0 chlamydia p

20	1408	72.1	402	2 Q46203	Q46203 chlamydia p
21	1407	72.1	402	2 Q46236	Q46236 chlamydia p
22	1382	70.8	402	2 Q9AIJ3	Q9AIJ3 chlamydia p
23	1371.5	70.3	352	2 Q70085	Q70085 chlamydia p
24	1367.5	70.1	352	2 Q70050	Q70050 chlamydia p
25	1363.5	69.9	352	2 Q69307	Q69307 chlamydia p
26	1362.5	69.8	352	2 Q69306	Q69306 chlamydia p
27	1357	69.5	337	2 Q9XBF6	Q9XBF6 chlamydophi
28	1353	69.3	336	2 Q9XBF5	Q9XBF5 chlamydophi
29	1353	69.3	353	2 Q69305	Q69305 chlamydia p
30	1352	69.3	353	2 Q9AIJ8	Q9AIJ8 chlamydia p
31	1351	69.2	404	2 Q46407	Q46407 chlamydia t
32	1343	68.8	389	2 Q9AIJ4	Q9AIJ4 chlamydia s
33	1342	68.8	386	2 Q69093	Q69093 chlamydia t
34	1341.5	68.7	396	2 Q9AIJ5	Q9AIJ5 chlamydia s
35	1339.5	68.6	397	2 Q69094	Q69094 chlamydia t
36	1337.5	68.5	397	2 Q69095	Q69095 chlamydia t
37	1336.5	68.5	387	2 Q9AIJ0	Q9AIJ0 chlamydia s
38	1335.5	68.4	397	2 Q9F951	Q9F951 chlamydia t
39	1333	68.3	340	2 Q9XBF2	Q9XBF2 chlamydophi
40	1331.5	68.2	387	2 Q9AIJ1	Q9AIJ1 chlamydia s
41	1331.5	68.2	397	2 Q9F950	Q9F950 chlamydia t
42	1330.5	68.2	385	2 Q9AIJ6	Q9AIJ6 chlamydia s
43	1328	68.0	396	2 Q46406	Q46406 chlamydia t
44	1327.5	68.0	385	2 Q9AIJ7	Q9AIJ7 chlamydia s
45	1326	67.9	356	2 Q52924	Q52924 chlamydia p

## ALIGNMENTS

RESULT 1

Q08085 ID Q08085 PRELIMINARY; PRT; 389 AA.  
AC Q08085;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP)  
OS Chlamydia psittaci (Chlamydophila psittaci).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
OX NCBI\_taxid=83554;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KOALA TYPE 1;  
RX MEDLINE=94171025; PubMed=8125292;  
RA Girjes A.A., Carrick F.N., Lavin M.F.;  
RT Remarkable sequence relatedness in the DNA encoding the major outer  
RT membrane protein of Chlamydia psittaci (koala type I) and Chlamydia  
RT pneumoniae.";  
RL Gene 138:139-142(1994).

CC !- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY  
CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH  
CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.  
CC !- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP  
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.  
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
DR EMBL; X72023; CAA50906.1; ;  
DR InterPro: IPR000604; Chlamydia\_OMP.  
DR Pfam: PF01308; Chlamydia\_OMP; 1.  
DR PRINTS; PR01334; CHLAMIDIACOMP.  
DR PRODOM; PD001717; Chlamydia\_OMP; 1.  
DR Outer membrane; Transmembrane; Porin; Signal.  
FT CHAIN 1 23 BY SIMILARITY.  
FT SIGNAL 24 389 MAJOR OUTER MEMBRANE PROTEIN.  
SQ SEQUENCE 389 AA; 41579 MW; 5DC50E85A6F4E50F CRC64;

Query Match 97.8%; Score 1910; DB 2; Length 389;  
Best Local Similarity 97.5%; Pred. No. 4.4e-147;  
Matches 357; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 2 LPVGNPSPDLLDGTWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61



```

|||||
Db 24 LPVGNPDPSSLIDGTIWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRLKLVDAKPTF 83
Qy 62 SMGAKPTGSAANYYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVDFCTLGASNG 121
Db 84 SMGAKPTGSAANYYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVDFCTLGASNG 143
Qy 122 YIRGNSTAFNLVGLFGYKGTNNANLNPVSLNGVVELYTDTSFSSVGARGALWECG 181
Db 144 YIKGNSTAFNLVGLFGYKGTNNANLNPVSLNGVVELYTDTSFSSVGARGALWECG 203
Qy 182 ATLGAEFQYQASPKPVEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATATGKTSAT 241
Db 204 ATLGAEFQYQASPKPVEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATATGKTSAT 263
Qy 242 NYHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301
Db 264 NYHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323
Qy 302 ATALSTTDSFDMQIVSQINKEFKSRKACGVTGATLVADADKWSLTAEARLINERAHV 361
Db 324 TTTLATSDSFDMQIVSQINKEFKSRKACGVTGATLVADADKWSLTAEARLINERAHV 383
Qy 362 SQGFRR 367
Db 384 SQGFRR 389

RESULT 2
Q9APM4
ID Q9APM4 PRELIMINARY; PRT; 389 AA.
AC Q9APM4; 2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMPL.
OS Chlamydia abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83555;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=LLG;
RX MEDLINE=20569239; PubMed=11119563;
RA Vreton E., Psarrou E., Kaisar M., Vlisidou I., Salti-Montesanto V.,
RT Longbottom D.;
RT "Identification of protective epitopes by sequencing of the major
RT outer membrane protein gene of a variant strain of Chlamydia psittaci
RL serotype 1.";
RL Infect. Immun. 69:607-612(2001).
DR EMBL: AF272945; AAG53881.1; -
KW Signal.
FT SIGNAL.
FT CHAIN 1 22 POTENTIAL.
FT CHAIN 23 389 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 389 AA; 41897 MW; 20513C69C7DBAA5 CRC64;

Query Match 78.5%; Score 1531.5; DB 2; Length 389;
Best Local Similarity 75.7%; Pred. No. 2.3e-116;
Matches 278; Conservative 38; Mismatches 50; Indels 1; Gaps 1;

Qy 2 LPVGNPDPSSLIDGTIWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRLKLVDAKPTF 61
Db 23 LPVGNPAEPSSLIDGTWMEGASGDDPCDPCSTWCDAISLRAGYGYGVDFRLKLVDNKTI 82
Qy 62 S-MGAKPTGSAANYYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVDFCTLGASN 120
Db 83 TGMGAVPTGTAADYKTPDRENPAYNKHLDHAEFTNAGFTALNIWDRFVDFCTLGASN 142
Qy 121 GYIRGNSTAFNLVGLFGYKGTNNANLNPVSLNGVVELYTDTSFSSVGARGALWECG 180
Db 143 GYFKASSAFAFLVGLFGYKGTNNANLNPVSLNGVVELYTDTSFSSVGARGALWECG 202

```

```

Qy 181 CATLGAEFQYQASPKPVEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATATGKTSAT 240
Db 203 CATLGAEFQYQASPKPVEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATATGKTSAT 262
Qy 241 INYHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLG 300
Db 263 IKYHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLG 322
Qy 301 NATALSTTDSFDMQIVSQINKEFKSRKACGVTGATLVADADKWSLTAEARLINERAHV 360
Db 323 EATLDTNNKFAFQLQIASIQINKMKSRKACGAVGATLIDADKWSITGEARLINERAHV 382
Qy 361 VSGQFRF 367
Db 383 MNAQFRF 389

RESULT 3
Q9AIK1
ID Q9AIK1 PRELIMINARY; PRT; 388 AA.
AC Q9AIK1; 2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VS225;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL: AF269259; AAK00240.1; -
KW Signal.
FT NON TER 1 1 POTENTIAL.
FT SIGNAL <1 19
FT CHAIN 20 388 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 388 AA; 41573 MW; 8E23D2C2C9B9948D CRC64;

Query Match 78.4%; Score 1529.5; DB 2; Length 388;
Best Local Similarity 76.2%; Pred. No. 3.3e-116;
Matches 281; Conservative 35; Mismatches 50; Indels 3; Gaps 2;

Qy 2 LPVGNPDPSSLIDGTIWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRLKLVDAKPTF 61
Db 20 LPVGNPAEPSSLIDGTWMEGASGDDPCDPCATWCDAISLRAGYGYGVDFRLKLVDNKTI 79
Qy 62 S-MGAKPTGSAANYYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVDFCTLGASN 120
Db 80 SGMGAAPTGSAAADYKTPDRENPAYNKHLDHAEFTNAGFTALNIWDRFVDFCTLGASN 139
Qy 121 GYIRGNSTAFNLVGLFGYKGTNNANLNPVSLNGVVELYTDTSFSSVGARGALWECG 180
Db 140 GYFKASSAFAFLVGLFGYKGTNNANLNPVSLNGVVELYTDTSFSSVGARGALWECG 199
Qy 181 CATLGAEFQYQASPKPVEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATATGKTSAT 238
Db 200 CATLGAEFQYQASPKPVEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATATGKTSAT 259
Qy 239 ATINYHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSL 298
Db 260 ATLYKHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSL 319
Qy 299 LCNATLSTTDSFDMQIVSQINKEFKSRKACGVTGATLVADADKWSLTAEARLINERAHV 358
Db 320 LGEATLSTTDSFDMQIVSQINKEFKSRKACGVTGATLVADADKWSLTAEARLINERAHV 379
Qy 359 AHVSGQFRF 367

```



```
Db 380 AHMNAQFRF 388
|||||

RESULT 4
Q46235
ID Q46235 PRELIMINARY; PRT; 391 AA.
AC Q46235;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN MOMP OR OMPA.
OS Chlamydia psittaci (Chlamydophila psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AVIAN TYPE C;
RA Storey C.C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GERMAN DUCK;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL; L25436; AAK0241.1; -.
DR EMBL; AF269261; AAK00242.1; -.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR PRINTS; PR01334; CHLAMIDIAOMP.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 391 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 391 AA; 42427 MW; D98A8F5E76362A00 CRC64;

Query Match 76.9%; Score 1501.5; DB 2; Length 391;
Best Local Similarity 73.2%; Pred. No. 6.2e-114;
Matches 270; Conservative 48; Mismatches 48; Indels 3; Gaps 2;

QY 2 LPVGNPSPSLIDGTWEGAGDPCDPCATWCDALISLRAGFYGYVDFRILKVDAPKTF 61
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 23 LPVGNPAEPLSLIDGTWEGAGDPCDPCSTWCDALISIRAGYGYDYVDFRVLKVDVNTKF 82
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 62 S-MGAKPTGSSAANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFDVFTLGASN 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 83 SGIGKPTGSSPNDFKNAEDRNVAIGHRLQDSEFTNAFLALNIWDRFDVFTLGASN 142
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 GYIRGNSTAFNLVGLFGVGTVTNANELPNVSLNSGVVELYDTDFSNVSGARGALWECG 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 143 GYFKASSAFLNLVGLGVKSSLTNDLPNVAITQGVVEFYDTDFSNVSGARGALWECG 202
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 CATLGAEFQYQSKPKVEELNVCNVSQFVNKPKYKQVA--FPLPTDAGVATATGTS 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 203 CATLGAEFQYQSKPKIEMNLNVISSPAQFVHKPRYKGTCTSANFPLPANAGTEAATDTKS 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 239 ATINYHEWQVGLSLVRLNSLYPYIGVQVSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 263 ATLKHWEQVGLSLVRLNMLVPYIGVNNSRATFDADTIRIAQPKLASAMNLTWNPTL 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 299 LGNATALSTDSFDMQIVSQINPKFKSRKACGVTVGATLVADKWSLTAEARLINERA 358
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 323 LGNATMLDTSNKFDFLQIASIQINKMKSRKACGLAIGATLADKWSITGEARLINERA 382
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 359 AHVSGQFRF 367
|||||:|||||
Db 383 AHMNAQFRF 391
|||||:|||||
```



DR	ENBL; AF269267; AAK00248.1;
KW	Signal.
FT	SIGNAL 1 22 POTENTIAL.
FT	CHAIN 23 392 MAJOR OUTER MEMBRANE PROTEIN.
SQ	SEQUENCE 392 AA; 42293 MW; FC31FC05195246C CRC64;

Query Match 76.2%; Score 1487; DB 2; Length 392;  
 Best Local Similarity 74.2%; Pred. No. 9.4e-113;  
 Matches 276; Conservative 39; Mismatches 49; Indels 8; Gaps

Qy	2	LPVGNPSDPSLLIDGTIWEGAGDCPCATWCDAISLRAGEYGDYVFDFRILKVDAPKTF 61
Dd	23	LPVGNPAEPSLLIDGTIWEGASGDCPCATWCDAISRAGYYGDYVDFRVILKVDNVNTF 82
Qy	62	SMGAK-----PTGSAAANYTTAVDRPNPAYNKKHLHDAEWTNAGFIALNIWDREFVFCITLG 117
Dd	83	SGMAKSPTETATGTASAT--TIVADRTNLAYCKKLQDAEWTNAAFALNIWDREFIFCTIG 141
Qy	118	ASNGYIRGNSHTAFNLGVLCFKVTNNANELPVLNSLUSGVVELYTDTSPFSWSVGANGALM 177
Dd	142	ASNGYFKASSAFAFNVLGLIKLKTDNF-NQLPNVAITQGVEFYDTTTFSSWSVGANGALM 200
Qy	178	ECCCATILGAEFYAQSQPKVEELNVICNVSQFSVPNKPKYGVA--FPILPTDAGVATATG 235
Dd	201	ECCCATILGAEFYAQSNPKIEMLNVTSSPAQFVIHKPRGYKGTSNFPPLPIDAGTDAATD 260
Qy	236	TKSATINHEWQVGSLSRYLNSLVPIYGVOWSRATFDADNIRIAOPKPLTAVLNLTAWN 295
Dd	261	TKSATLYKHQVGLALSRYLNMVPIYGVNWSRAFDADTRIAPQKPLATAVIDLTTWN 320
Qy	296	PSSLGNATALSTTDSFSDFMQIVSCQINKPKSKACGVTVGATLYDADKWSLTAEARLIN 355
Dd	321	PTLLGKATTVDGINTYSDFIQLASIQINKMKSRKACGVAVGATLIDADKWSITGEARLIN 380
Qy	356	ERAAHVSGQPRF 367
Dd	381	ERAAHMNAQPRF 392

RESULT	8
Q9AI12	
ID	Q9AI12 PRELIMINARY; PRT; 381 AA.
AC	Q9AI12
DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN	OMPA.
OS	Chlamydomophila pecorum.
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX	NCBI_TaxID=85991;
RN	[1]
RP	SEQUENCE OF 34-369 FROM N.A.
RC	STRAIN=1710S;
RX	MEDLINE=93123168; PubMed=8419295;
RA	Kaltenboeck B., Kousoulas K.G., Storz J.;
RT	"Structures of and allelic diversity and relationships among the major outer membrane protein (ompA) genes of the four chlamydial species";
RL	J. Bacteriol. 175:487-502(1993).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=1710S;
RX	MEDLINE=21078680; PubMed=11211261;
RA	Bush R.M., Everett K.D.;
RT	"Molecular evolution of the Chlamydiaceae.";
RL	Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=1710S;
RA	Everett K.D.E., Hamblly W.A., Andersen A.A.;
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR	ENBL; AF269279; AAK00260.1; -







QY 237 KSAITNHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNP 296  
 Db 262 KSAITVKEWQVGLSYRLNMLVPYIGVQWSRATFDADNIRIAQPKLASAILNLTWNP 321  
 QY 297 SLLGNATSLTDSFDFMQIVSQINKFKSRKACGVTVGATLVADKWSLTAEARLINE 356  
 Db 322 TLLGVATTLTNSKYADFQIVSQINKFKSRKACGVTVGATLVADKWSLTAEARLINE 381  
 QY 357 RAAHVSQGRF 367  
 Db 382 RAAHINAQRF 392

RESULT 11  
 Q9AIII PRELIMINARY; PRT; 380 AA.  
 AC Q9AIII;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).  
 GN OMPA.  
 OS Chlamydomonas reinhardtii (Chlamydomonadales; Chlamydomonadales; Chlamydomonadales).  
 OC Bacteria; Chlamydomonadales; Chlamydomonadales; Chlamydomonadales.  
 OX NCBI\_TaxID=85991;  
 RN [1]  
 RP SEQUENCE OF 40-352 FROM N.A.  
 RC STRAIN=L71;  
 RX MEDLINE=93123168; PubMed=8419295;  
 RA Kallenboeck B., Kousoulas K.G., Storz J.;  
 RT "Structures of and allelic diversity and relationships among the major  
 RT outer membrane protein (OmpA) genes of the four chlamydial species."  
 RL J. Bacteriol. 175:487-502(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=L71;  
 RX MEDLINE=21078680; PubMed=11211261;  
 RA Bush R.M., Everett K.D.;  
 RT "Molecular evolution of the Chlamydiae."  
 RT Int. J. Syst. Evol. Microbiol. 51:203-220(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=L71;  
 RX Everett K.D.E., Hamby W.A., Andersen A.A.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF269280; AA00261.1;  
 KW Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL <1 15  
 FT CHAIN 16 380  
 FT MAJOR OUTER MEMBRANE PROTEIN.  
 SQ SEQUENCE 380 AA; 41047 MW; 829A18D3C5A85008 CRC64;

Query Match 75.3%; Score 1469.5; DB 2; Length 380;  
 Best Local Similarity 71.4%; Pred. No. 2.4e-111;  
 Matches 262; Conservative 54; Mismatches 48; Indels 3; Gaps 3;  
 QY 2 LPVGNPSPDLLIDGTWEGAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 Db 16 LPVGNPSPDLLIDGTWEGAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 75  
 QY 62 SNGAKPTGSA--AANYTTAVDRENPAYNKHLDHDAEFTNAGFTALNIWDRFDVCTLGAS 120  
 Db 76 LMGTAFTSPNSAASNTTAEARNPAYGKHMDHDAEFTNAGFTALNIWDRFDVCTLGAS 135  
 QY 121 GYIRGNSTAFNLVGLFGVKGTTVNAELPNVLSNGVVELYTDTSFSSVSGARGALWECG 180  
 Db 136 GYFGKNSAFNLIGLIGSGTLD--QKYPNASISNGVVELYTDTSFSSVSGARGALWECG 194  
 QY 181 CATLGAEEFYAQSKPKVEELNVCNVSQFVSNKPKYKGVAFPLPTDAGVATATGTSAT 240  
 Db 195 CATLGAEEFYAQSKPKRVDELNVLNVAQFTVHKFGYVQSGSLPLPTNAGTSNASDLKNT 254

QY 241 INHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNP 300  
 Db 255 INHEWQVGAALSRYRLNMLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNP 314  
 QY 301 NATALSTDSFDFMQIVSQINKFKSRKACGVTVGATLVADKWSLTAEARLINE 360  
 Db 315 QATSVG--SDRYADTLQIVSLQINKFKSRKACGVSMGATLLDADKWSLTAEARLINE 373  
 QY 361 VSGQRF 367  
 Db 374 LSAQCRF 380

RESULT 12  
 Q9AIJ2 PRELIMINARY; PRT; 391 AA.  
 AC Q9AIJ2;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).  
 GN OMPA.  
 OS Chlamydia psittaci (Chlamydiales; Chlamydiales; Chlamydiales).  
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
 OX NCBI\_TaxID=83554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WC;  
 RX MEDLINE=21078680; PubMed=11211261;  
 RA Bush R.M., Everett K.D.;  
 RT "Molecular evolution of the Chlamydiae."  
 RT Int. J. Syst. Evol. Microbiol. 51:203-220(2001).  
 DR EMBL; AF269269; AA00250.1;  
 KW Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL <1 17  
 FT CHAIN 18 391  
 FT MAJOR OUTER MEMBRANE PROTEIN.  
 SQ SEQUENCE 391 AA; 42056 MW; B58A73D4980408BF CRC64;

Query Match 72.9%; Score 1423; DB 2; Length 391;  
 Best Local Similarity 71.1%; Pred. No. 1.5e-107;  
 Matches 266; Conservative 44; Mismatches 56; Indels 8; Gaps 5;  
 QY 2 LPVGNPSPDLLIDGTWEGAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61  
 Db 18 LPVGNPSPDLLIDGTWEGAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 77  
 QY 62 S-MGAKPTGSA--AANYTTAVDRENPAYNKHLDHDAEFTNAGFTALNIWDRFDVCTLGA 118  
 Db 78 SGMAAIPTESSGTVSSAKQAVDRVNLAYGKHLQDAEFTNAGFTALNIWDRFDVCTLGA 137  
 QY 119 SNGYIRGNSTAFNLVGLFGVKGTT--VNA--NELPNVLSNGVVELYTDTSFSSVSGARGA 175  
 Db 138 SNGYFKGSSAFLNVLGFLGFIAGNSNALNDQLPNAVITQIVFEYTDTSFSSVSGARGA 197  
 QY 176 LWEGCATLGAEEFYAQSKPKVEELNVCNVSQFVSNKPKYKGVAFPLPTDAGVATA 233  
 Db 198 LWEGCATLGAEEFYAQSKPKVEELNVCNVSQFVSNKPKYKGVAFPLPTDAGVATA 257  
 QY 234 TGTSATINHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLT 293  
 Db 258 TDKSATIKYHEWQVGLSYRLNMLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLT 317  
 QY 294 WNPILLGNATSLTDSFDFMQIVSQINKFKSRKACGVTVGATLVADKWSLTAEARL 353  
 Db 318 WNPITGASGVSDNTNKNWSDMLQIASIQINKKSRKACGVAVGATLVADKWSLTAEARL 377  
 QY 354 INERAHVSGQRF 367  
 Db 378 INERAHVNAQRF 391



```

RESULT 13
Q9X17 PRELIMINARY; PRT; 341 AA.
AC Q9X17;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydomydia abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.
OX NCBI_TaxID=83555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=I4508;
RX MEDLINE=93123168; PubMed=8419295;
RA Kallenberg B., Kousoulas K.G., Storz J.;
RT "Structures of and allelic diversity and relationships among the major
RL outer membrane protein (ompA) genes of the four chlamydial species.";
DR EMBL; M73040; AAD29103.1;
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
FT NON_TER 1
FT CHAIN 341
SQ SEQUENCE 341 AA; 36762 MW; B5933C9BF6AAFL171 CRC64;

Query Match 72.8%; Score 1420.5; DB 2; Length 341;
Best Local Similarity 75.4%; Pred. No. 1.9e-107;
Matches 257; Conservative 35; Mismatches 48; Indels 1; Gaps 1;

QY 16 GTWGAAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDAPKTF--MGAKPTGSAAN 74
Db 1 GTWGAAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDAPKTF--MGAKPTGSAAN 74

QY 75 YTTAVDRPNPANKHLHDAEFTNAGFIALNIWDRFVCTLGASNGYIRGNSTAFNLV 134
Db 1 YTTAVDRPNPANKHLHDAEFTNAGFIALNIWDRFVCTLGASNGYIRGNSTAFNLV 134

QY 61 YKPTDRPNPANKHLHDAEFTNAGFIALNIWDRFVCTLGASNGYIRGNSTAFNLV 120
Db 1 YKPTDRPNPANKHLHDAEFTNAGFIALNIWDRFVCTLGASNGYIRGNSTAFNLV 120

QY 135 LFCVKGTNNANLPNLSNGVVELYDTSTFWSVGARGALWECGCATLGAEFYAQS 194
Db 1 LFCVKGTNNANLPNLSNGVVELYDTSTFWSVGARGALWECGCATLGAEFYAQS 194

QY 121 LGVKGSSAADQLPNWGTQGVFEYDTSTFWSVGARGALWECGCATLGAEFYAQS 180
Db 1 LGVKGSSAADQLPNWGTQGVFEYDTSTFWSVGARGALWECGCATLGAEFYAQS 180

QY 195 PKVEELNVCNVSQFVNPKGKVAFLPTDAGVATATGKATINHEWQVGLASY 254
Db 1 PKVEELNVCNVSQFVNPKGKVAFLPTDAGVATATGKATINHEWQVGLASY 254

QY 181 PKIEMLVNVSPPAQFVHPRGKGTAFPLTAGTDQATDYKATIKYHEWQVGLASY 240
Db 1 PKIEMLVNVSPPAQFVHPRGKGTAFPLTAGTDQATDYKATIKYHEWQVGLASY 240

QY 255 RLNSLPYIGVQNSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLSTDSFDF 314
Db 1 RLNSLPYIGVQNSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLSTDSFDF 314

QY 241 RLNLVPIYGVNWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLSTDSFDF 300
Db 1 RLNLVPIYGVNWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLSTDSFDF 300

QY 315 MQIVSCQINKSKRACGVTGATLVADKWSLTAEARLIN 355
Db 1 MQIVSCQINKSKRACGVTGATLVADKWSLTAEARLIN 355

QY 301 LQIASIQINKSKRACGVTGATLVADKWSLTAEARLIN 341
Db 1 LQIASIQINKSKRACGVTGATLVADKWSLTAEARLIN 341

RESULT 14
Q46193 PRELIMINARY; PRT; 402 AA.
AC Q46193;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMP.
OS Chlamydia psittaci (Chlamydomydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=94103736; PubMed=8277245;
RA Storey C., Lusher M., Yates P., Richmond S.;
RT "Evidence for Chlamydia pneumoniae of non-human origin.";
RL J. Gen. Microbiol. 139:2621-2626(1993).
DR EMBL; L04980; AAA17396.1;
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR PRINTS; PR01334; CHLAMIDIAOMP.
DR ProDom; PD001717; Chlamydia_OMP; 1.
FT SIGNAL 1
FT CHAIN 23
FT CHAIN 402
SQ SEQUENCE 402 AA; 43261 MW; E36ABC5AF04820A CRC64;

Query Match 72.7%; Score 1419; DB 2; Length 402;
Best Local Similarity 70.5%; Pred. No. 3.2e-107;
Matches 268; Conservative 37; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSDPSLLIDGTWEGAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDAPKTF 61
Db 2 LPVGNPSDPSLLIDGTWEGAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDAPKTF 61

QY 23 LPVGNPAEPLSLIDGTWEGAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDAPKTF 82
Db 23 LPVGNPAEPLSLIDGTWEGAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDAPKTF 82

QY 62 S-MGAKP---TGSA--ANYTTAVDRPNPANKHLHDAEFTNAGFIALNIWDRFVCTL 116
Db 62 S-MGAKP---TGSA--ANYTTAVDRPNPANKHLHDAEFTNAGFIALNIWDRFVCTL 116

QY 83 SGMAATPTQATGNASNTNQPEANGRENIAAYGRHMDAEWFSNAFLALNIWDRFVCTL 142
Db 83 SGMAATPTQATGNASNTNQPEANGRENIAAYGRHMDAEWFSNAFLALNIWDRFVCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKGITVNAE---LPNVLSNGVVELYDTSTFWSVGA 172
Db 117 GASNGYIRGNSTAFNLVGLFGVKGITVNAE---LPNVLSNGVVELYDTSTFWSVGA 172

QY 143 GASNGYIRGNSTAFNLVGLFGVKGITVNAE---LPNVLSNGVVELYDTSTFWSVGA 202
Db 143 GASNGYIRGNSTAFNLVGLFGVKGITVNAE---LPNVLSNGVVELYDTSTFWSVGA 202

QY 173 RGALWECGCATLGAEFYAQSAPKVEELNVCNVSQFVNPKGKVA--FPLPTDAGV 230
Db 173 RGALWECGCATLGAEFYAQSAPKVEELNVCNVSQFVNPKGKVA--FPLPTDAGV 230

QY 203 RGALWECGCATLGAEFYAQSAPKVEELNVCNVSQFVNPKGKVA--FPLPTDAGV 262
Db 203 RGALWECGCATLGAEFYAQSAPKVEELNVCNVSQFVNPKGKVA--FPLPTDAGV 262

QY 231 ATATGKTSATINHEWQVGLASYRLNSLPYIGVQNSRATFDADNIRIAQPKLPTAVLN 290
Db 231 ATATGKTSATINHEWQVGLASYRLNSLPYIGVQNSRATFDADNIRIAQPKLPTAVLN 290

QY 263 TTTATGKTSATINHEWQVGLASYRLNSLPYIGVQNSRATFDADNIRIAQPKLPTAVLN 322
Db 263 TTTATGKTSATINHEWQVGLASYRLNSLPYIGVQNSRATFDADNIRIAQPKLPTAVLN 322

QY 291 LPAWNPSSLGNATLSTT---DSFDEMOIVSCQINKSKRACGVTGATLVADKWSL 347
Db 291 LPAWNPSSLGNATLSTT---DSFDEMOIVSCQINKSKRACGVTGATLVADKWSL 347

QY 323 ITWNPSSLGCTALPNNAKGVLDVLIQIASIQINKSKRACGVTGATLVADKWSL 382
Db 323 ITWNPSSLGCTALPNNAKGVLDVLIQIASIQINKSKRACGVTGATLVADKWSL 382

QY 348 TAEARLINERAAHVSGQFRF 367
Db 348 TAEARLINERAAHVSGQFRF 367

QY 383 TGEARLINERAAHVSGQFRF 402
Db 383 TGEARLINERAAHVSGQFRF 402

RESULT 15
Q9AIJ9 PRELIMINARY; PRT; 382 AA.
AC Q9AIJ9;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydomydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MENINGOPNEUMONITIS, MN, ATCC VR122;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL; AF269262; AAK00243.1;
FT SIGNAL 1
FT CHAIN 2
FT CHAIN 382
SQ SEQUENCE 382 AA; 41231 MW; 6917171A9A69303B CRC64;

```



```

Query Match      72.5%; Score 1415; DB 2; Length 382;
Best Local Similarity 70.3%; Pred. No. 6.3e-107;
Matches 267; Conservative 37; Mismatches 62; Indels 14; Gaps 6;

QY 2 LPVGNPSLLIDGTWEGAGDPCDPCATWCDAISIRAGFYGDYVDFDRILKVDAPKTF 61
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 LPVGNPAEPSLLIDGTWEGASGDCDPCATWCDAISIRAGYGYGDYVDFRVLKVDNKT 62
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 62 S-MGAKP---TGSAA-ANYTTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRDFVCT 116
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 63 SGMAATPTQATGNASNTNQPEANGRPNTAYGRHMDAEWFSNAAFLALNIWDRDFICT 122
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 117 GASNGYIRCNSTAFNLVGLFGVKGTTVNANE---LPNVSLNGVVVELYTDTSFS 172
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 123 GASNGYFKSSSAFNLVGLIGFSATSTSTELPMQLPNVGTQGVVEFTDTSFSWS 182
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 173 RGALWECGCATLGAEFQYQAQSKPKVEELNVCNVSQFSVKNPKYKGVA--FPLPTDAGV 230
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 183 RGALWECGCATLGAEFQYQAQSNPKIEVLNVTSSPAQFVIHKPRGYKGASSNEPLITAGT 242
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 231 ATATGTSATINYNHQQWQASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 243 TEATDTKSATIKYHEWQVGLALSRYRLNMLVPYIGVNWRSRATFDADTIRIAQPKLSEILN 302
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 291 LTAWNPSLLGNATALSTT--DSFSDQMIVSCQINKEKSKACGVTVCATLVDADKWSL 347
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 303 ITTWNPSLLGSTTTLPNNGGKDVLSDLVQIASIQINKMKSKACGVAVGATLIDADKWSI 362
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 348 TAEARLINERAAHVSGQFRF 367
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 363 TGEARLINERAAHMAQFRF 382
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Search completed: February 7, 2002, 21:41:22  
Job time: 4651 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using .sw model

Run On: February 7, 2002, 15:48:41 ; Search time 137.02 Seconds  
(without alignments)  
117.311 Million cell updates/sec

Title: US-09-391-606-7

Perfect score: 1166

Sequence: 1 MTKKHVAVVVEGILNRLPKQ.....DKLGSDFTFRKFDLGIISAF 217

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1162	99.7	422	21	AA192716 Chlamydia pneumoniae
2	1162	99.7	438	20	AA193539 Chlamydia pneumoniae
3	765.5	65.7	372	20	AA193750 Chlamydia trachoma
4	88	7.5	951	20	AA193436 Porphyromonas ging
5	88	7.5	953	20	AA193403 Porphyromonas ging
6	83	7.1	711	19	AA1955103 Porphyromonas ging
7	83	7.1	2234	21	AA1981502 Streptococcus pneu
8	81	6.9	1091	21	AA193631 Streptococcus pneu
9	80	6.9	303	22	AA1982023 Amino acid sequenc
10	80	6.9	359	22	AA1982824 S. epidermidis ope
11	79	6.8	327	20	AA197217 H. pylori outer me

12	78	6.7	753	21	AA196159
13	77.5	6.6	2001	18	AA193596
14	77.5	6.6	3672	18	AA1931950
15	77.5	6.6	3801	18	AA1931949
16	77	6.6	500	19	AA193957
17	76.5	6.6	367	20	AA194087
18	76.5	6.6	367	21	AA198457
19	76.5	6.6	725	20	AA198832
20	75.5	6.5	431	20	AA195534
21	75.5	6.5	1329	21	AA198831
22	75	6.4	1329	21	AA1981779
23	74.5	6.4	241	19	AA195085
24	74.5	6.4	652	22	AA1966046
25	74.5	6.4	730	22	AA1966045
26	74.5	6.4	761	21	AA194084
27	74.5	6.4	761	21	AA199418
28	74.5	6.4	761	22	AA1966043
29	74.5	6.4	761	22	AA1966167
30	73.5	6.3	988	13	AA194302
31	73	6.3	415	21	AA191576
32	73	6.3	476	21	AA1930604
33	72.5	6.2	459	20	AA196942
34	72.5	6.2	561	18	AA198311
35	72.5	6.2	608	22	AA194640
36	72.5	6.2	762	21	AA1937984
37	72	6.2	405	22	AA19943
38	72	6.2	701	22	AA192442
39	71.5	6.1	930	19	AA195961
40	71.5	6.1	1395	20	AA193563
41	71.5	6.1	1395	20	AA198401
42	71	6.1	288	22	AA193621
43	71	6.1	354	21	AA1929125
44	70.5	6.0	112	21	AA192299
45	70.5	6.0	112	21	AA1945000

#### ALIGNMENTS

RESULT 1  
AA192716  
ID AA192716 standard; Protein: 422 AA.  
AC AA192716;  
DT 29-AUG-2000 (first entry)  
DE Chlamydia pneumoniae lorf2.  
KW Lorf2; vaccine; antibacterial; antigen.  
OS Chlamydia pneumoniae.  
PN WO200024901-A1.  
PD 04-MAY-2000.  
PF 28-OCT-1999; 99WO-GB03565.  
PR 28-OCT-1998; 98US-0106037.  
PR 20-SEP-1999; 99US-0154658.  
PR 26-OCT-1999; 99US-0427501.  
(CONN-) CONNAUGHT LAB LTD.  
Murdin AD, Oomen RP, Dunn PL;  
WPI; 2000-350742/30.  
N-PSDB; AAA28411.

Isolated polynucleotide encoding a Chlamydia polypeptide useful to treat, diagnose and prevent disease caused by Chlamydia infection



Claim 6; Fig 1A-C; 88pp; English.

PS This is the lorf2 protein of a strain of Chlamydia pneumoniae.  
 CC Comparison of this sequence as to the recently published genome sequence  
 CC of C. pneumoniae reveals that the sequence actually contains at least two  
 CC open reading frames, a first one in the 5' portion and a second one in  
 CC the 3' portion of the sequence. Despite the presence of the stop codon at  
 CC the end of this sequence, C. pneumoniae does make a 76 kDa product. It  
 CC appears possible that C. pneumoniae is able to read through this stop  
 CC codon and produce a full-length product terminated by the stop codon at  
 CC the end of the second open reading frame. There is at least one in-frame  
 CC ATG upstream of the start codon. This suggests that the first open  
 CC reading frame may form part of one or more larger open reading frames.  
 CC The lorf2 protein or DNA can be used as a vaccine for humans to treat or  
 CC prevent disease caused by Chlamydia infection. The sequences or an  
 CC antibody to lorf2 can be used to diagnose a Chlamydial infection.

SQ Sequence 422 AA;

Query Match 99.7%; Score 1162; DB 21; Length 422;  
 Best Local Similarity 99.5%; Pred. No. 4.8e-119;  
 Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAYWVEGILNRLPKQFFVKCSVDWNTFVPSSTTTEKAATNAMKYKVCWQWLV 60  
 DB 206 mtkkhyawvvegilnrlpkqffvkcsvvdwntfvpststtekaatnamkykvcwqwl 265  
 QY 61 GKHSQVPWINGOKKPLYLYGAFLMNPLAKATKTTLNGKENLAWFTGGTGLGRLKAGDWSA 120  
 DB 266 gkhsqvpwngokkplylygafilmplakatkttlنگkenlawfiggtlglrkagdw 325  
 QY 121 TVRYEYVEALSVPETDVSIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYMYGI 180  
 DB 326 tvryeyvealsvpelidvsgirgnllkfafaqaiaanydpkeangftnykgsalymygi 385  
 QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217  
 DB 386 tdslsfraygayskpandksgdftfrkfdlgiisaf 422

RESULT 2

AA35359  
 ID AAY35359 standard; Protein; 438 AA.

AC AAY35359;

DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae transmembrane protein sequence.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN WO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GEST ) GENSET.

XX Griffais R;

DR WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

XX

PS Page 1157; Disclosure; 1912pp; English.

CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAY34584-Y35879) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.

SQ Sequence 438 AA;

Query Match 99.7%; Score 1162; DB 20; Length 438;  
 Best Local Similarity 99.5%; Pred. No. 5.1e-119;  
 Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAYWVEGILNRLPKQFFVKCSVDWNTFVPSSTTTEKAATNAMKYKVCWQWLV 60  
 DB 222 mtkkhyawvvegilnrlpkqffvkcsvvdwntfvpststtekaatnamkykvcwqwl 281  
 QY 61 GKHSQVPWINGOKKPLYLYGAFLMNPLAKATKTTLNGKENLAWFTGGTGLGRLKAGDWSA 120  
 DB 282 gkhsqvpwngokkplylygafilmplakatkttlنگkenlawfiggtlglrkagdw 341  
 QY 121 TVRYEYVEALSVPETDVSIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYMYGI 180  
 DB 342 tvryeyvealsvpelidvsgirgnllkfafaqaiaanydpkeangftnykgsalymygi 401  
 QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217  
 DB 402 tdslsfraygayskpandksgdftfrkfdlgiisaf 438

RESULT 3

AA37570  
 ID AAY37570 standard; Protein; 372 AA.

AC AAY37570;

DT 07-OCT-1999 (first entry)

DE Chlamydia trachomatis cellular envelope protein.

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perinephritis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

PN WO9928475-A2.

PD 10-JUN-1999.

PF 27-NOV-1998; 98WO-IB01939.

PR 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97FR-0015041.

PR 17-DEC-1997; 97FR-0016034.

XX (GEST ) GENSET.

XX Griffais R;

DR WPI; 1999-371125/31.

PT Genome sequence of Chlamydia trachomatis







XX WPI: 1999-385613/32.  
DR N-PSDB; AAX91621.  
XX Antigenic Porphyromonas gingivalis peptides for preventing  
PT gingivitis  
XX  
XX Claim 1; Page 373-375; 588pp; English.  
XX  
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to  
CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the  
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
CC activity with a vaccine mechanism of action. The PG polypeptides can be  
CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
CC be used to detect Porphyromonas gingivalis in standard hybridisation  
CC assays. Porphyromonas gingivalis is involved in periodontal disease  
CC especially gingivitis.  
XX  
XX Sequence 953 AA;  
SQ

Query Match 7.5%; Score 88; DB 20; Length 953;  
Best Local Similarity 30.1%; Pred. No. 1.4;  
Matches 25; Conservative 17; Mismatches 35; Indels 6; Gaps 4;  
QY 117 DWSATVRYEYVLSVPEIDVSGIGRG-NLLKFWFAQIAAANYDPKEAN---SFTNYKGF 172  
DB 497 dwkngmrhsvpsltvplldylnltmgvnynewytkgrkswnedkktflpsdttk-f 555  
QY 173 SALLYWGITDSLSFRAYGAYSKP 195  
DB 556 rrllydsiaagistltlygmf-kp 577  
RESULT 6  
ID AAW55103 standard; Protein; 711 AA.  
XX  
AC AAW55103;  
DT  
DT 02-OCT-1998 (first entry)  
XX Streptococcus pneumoniae SP0051 protein.  
DE  
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
KW detection; pneumonia; otitis media; meningitis.  
XX Streptococcus pneumoniae.  
OS  
PH Key Location/Qualifiers  
FT Misc-difference 56 /label= unknown  
FT /note= "encoded by NAG"  
FT  
XX WO9818930-A2.  
XX  
XX 07-MAY-1998.  
XX  
XX 30-OCT-1997; 97WO-US19422.  
XX  
XX 31-OCT-1996; 96US-0029960.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;  
XX  
XX WPI: 1998-272224/24.  
DR N-PSDB; AAX27364.  
XX  
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus  
PT pneumoniae - or their epitope-containing fragments, useful in  
PT protective or therapeutic vaccines, and for diagnosis  
PT

XX Claim 11; Page 66; 118pp; English.  
XX  
XX The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
XX (especially 10-300) mu g/ml per dose.  
XX  
XX Sequence 711 AA;  
SQ

Query Match 7.1%; Score 83; DB 19; Length 711;  
Best Local Similarity 23.4%; Pred. No. 3.2;  
Matches 57; Conservative 28; Mismatches 81; Indels 78; Gaps 14;  
QY 32 TFVPSSTTTEKAATNAMYKYCVWOLYKHSQVPWINGQKPKLYLYGAFI----MNPL 87  
DB 443 thnpaseqtliqaaeigl-----lvqeeafdtwygg-kkp-ydygrffekdathpe 491  
QY 88 AKA-----TKTTL-NGKENLA---WFTGTLGLLRKAGDWSATVRYEYVLSVPEI 135  
DB 492 arkgekwsdflrtmvergknpaifmwsigneigangdahslatvk-rlvkv--ikdv 548  
QY 136 DVSGIGRGNLLKFWEA-----QAIANYPDKANSFTNYKGFSAIY----MY 178  
DB 549 dktryvtmgadkfrfgngsgghekiadeldavgfny-----sednykalrakhpkwliy 602  
QY 179 GITDSLSFRAYGAYSKP-----ANDKLG-----SDTFERKFDLGI 213  
DB 603 gsetssatrtgryrperelkhsngpernyeqsdyngdrvgwgtatstfdnag 662  
QY 214 ISAF 217  
DB 663 agqf 666  
RESULT 7  
ID AAY81502 standard; Protein; 2234 AA.  
XX  
AC AAY81502;  
DT  
DT 24-MAY-2000 (first entry)  
XX Streptococcus pneumoniae type 4 protein sequence #2.  
DE  
XX Streptococcus pneumoniae; vaccine; screening; protein antigen;  
KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
KW pneumococcal disease.  
XX Streptococcus pneumoniae.  
OS  
XX WO200006737-A2.  
XX  
XX 10-FEB-2000.  
XX  
XX 27-JUL-1999; 99WO-GB02451.  
XX  
XX 27-JUL-1998; 98GB-0016337.  
XX 19-MAR-1999; 99US-0125164.  
XX  
XX (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
XX



PI	Gilbert CFG, Hansbro PM;
XX	
DR	WPI; 2000-195300/17.
XX	
PT	New Streptococcal protein, useful as a vaccine, for diagnosis of
PT	pneumococcal diseases and for screening agents capable of antagonizing
XX	or inhibiting expression of the protein -
XX	
XX	Claim 1; Page 72; 108pp; English.
XX	
CC	AA051501 to AA051679 represent specifically claimed protein sequences
CC	isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent
CC	specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC	The sequences have antibacterial and antiinflammatory properties.
CC	The protein sequences, and fragments of them, are useful as immunogens
CC	and/or antigens. The nucleotide sequences can be used in vaccines and in
CC	diagnostic assays. The proteins and nucleotides can be useful for the
CC	detection and diagnosis of S. pneumoniae. The protein sequences are also
CC	useful for screening an agent capable of antagonising, inhibiting or
CC	interfering with the function or expression of the proteins in which the
CC	agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC	and meningitis. AA05591 to AA05614 represent primers used in the
CC	exemplification of the present invention.
XX	
XX	Sequence 2234 AA;
SQ	

```

Query Match      7.1%  Score 83;  DB 21;  Length 2234;
Best Local Similarity 23.4%;  Pred. No. 16;
Matches 57;  Conservative 28;  Mismatches 81;  Indels 78;  Gaps 14;

Qy 32  TFVPSETSTTEKAATNAWKYKVCWQMLVKGKHSQVPWINGOKKPLXLYLGAFL-----MNPL 87
   | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db 483 thnpasetiqliaaelg|-----lvqeeafdtwygg-kkp-ydygrfrefkdathpe 531

Qy 88  AKA-----TKTTL-NGKENLA---WFIGTGLGGLRKAGDWSATRYRYVEALSVEPI 135
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 532 arkgekwsdfolrlmvergknnpaifmwsagneigangdahslatvk-rlvky---ikdv 588

Qy 136 DVSGIGRGNLLKWFKA-----QAIAANYDPKEANSTNTYKGSFALY-----MY 178
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 589 dktryvmgadkfzfgnpgsgghekiadelavgfny-----sednykalrakthpkwliy 642

Qy 179 GITDLSLFRAYGAYSKP-----ANDKLG-----SDFTFRKFDLGI 213
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 643 gsctsatlrtrgsyrrrelkhsngpernyeqsdyndrvvgwgdktaswffdrdnag 702

Qy 214 ISAF 217
   |
Db 703 aaqf 706

```

RESULT	8
AAV43631	
ID	AAV43631 standard; Protein; 1091 AA.
XX	
XX	AAV43631;
XX	
DT	11-FEB-2000 (first entry)

XX	hmc gene; 3-hydroxy-3-methylglutaryl-CoA synthase; HMG-CoA synthase;
XX	mevalonate pathway; carotenogenic yeast; isopentenyl pyrophosphate;
KW	farnesyl pyrophosphate; isoprenoid; carotenoid; astaxanthin;
KW	cancer; antioxidant; colouring reagent; farmed fish industry.
XX	
XX	Phaffia rhodozyma.
XX	
XX	OS
XX	PN
XX	EP955363-A2.
XX	
XX	10-NOV-1999.
XX	
XX	

PF	26-APR-1999;	99EP-0107413.	
XX			
XX	06-MAY-1998;	98EP-0108210.	
PR			
XX			
PA	(HOFF ) HOFFMANN LA ROCHE & CO AG F.		
XX			
XX	Hoshino T, Ojima K, Setoguchi Y;		
XX			
DR	WPI: 2000-001086/01.		
DR	N-PSDB; AAZ30171.		
XX			
PT	Isolated DNA sequences encoding enzymes, useful for the production of		
PT	isoprenoids and carotenoids		
XX			
XX	Claim 3; Page 37-40; 58pp; English.		
PS			
XX			
CC	The present sequence is represents a 3-hydroxy-3-methylglutaryl-CoA		
CC	synthase (HMG-CoA synthase) enzyme, and is encoded by the hmc gene.		
CC	The enzyme is involved in the mevalonate pathway in the carotenogenic		
CC	yeast <i>Phaffia rhodozyma</i> . The specification also describes enzymes that		
CC	are involved in the pathway from isopentenyl pyrophosphate to farnesyl		
CC	pyrophosphate. The enzymes of the invention are used in the production		
CC	of isoprenoids and carotenoids, especially astaxanthin. Astaxanthin is		
CC	useful for the pharmaceutical industry, to protect cells against cancer		
CC	as it has a strong antioxidation property. Astaxanthin is also useful		
CC	as a colouring reagent in the farmed fish industry, e.g. salmon.		
XX			
XX	Sequence 1091 AA;		

```

Query Match      6.9%; Score 81; DB 21; Length 1091;
Best Local Similarity 22.4%; Pred. No. 9.8;
Matches 48; Conservative 30; Mismatches 50; Indels 86; Gaps 12;

QY      8 WV--VEGI-----LNRLPKQFFVKCSVDWNTFVPSST-----TEKA 44
      1111: : : : : : : : : : : : : : : : : : : : : : : : :
Db      756 wldsvegmevmaafnstfsfarlqskcmagrslyrlatstgdammnmagkgteka 815
      1111: : : : : : : : : : : : : : : : : : : : : : : : :

QY      45 ATNAMY-----KYCVQWMLVKKHSQVPWINGQKKPL-----YLXGAFLMNP LAK 89
      : : : : : : : : : : : : : : : : : : : : : : : : :
Db      816 letlseyfsmqlhalsgnycidk---kpsainwlegrgksvvaesvlpaiaksvl-- 869

QY      90 ATKYTT-----LNGKENLAWFIQ---GTJLGLLR-----KAGDWSATV 122
      1111: : : : : : : : : : : : : : : : : : : : : : : : :
Db      870 --kttvadlnlnhniknl---igsamagsiggfnoahasdilsflatgqdpagqvessm 924

QY      123 RYEVREAL-----SVPEIDVSGIGRGNLL 146
      : : : : : : : : : : : : : : : : : : : : : : : : :
Db      925 cmtlmeavndgkdllitcsmipaieqcvtvqqgtfl 958
      : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	9
AAG82023	
ID	AAG82023 standard; Protein; 303 AA.
XX	
AC	AAG82023;
XX	
DT	03-SEP-2001 (first entry)
XX	
DE	S. epidermidis open reading frame protein sequence SEQ ID NO:1140.
XX	
KW	Staphylococcus epidermidis Srl strain; infection; diagnosis;
KW	vaccination; endocarditis.
XX	
OS	Staphylococcus epidermidis.
XX	
PN	WO200134809-A2.
XX	
PD	17-MAY-2001.
XX	
PF	09-NOV-2000; 2000WO-US30782.
XX	
PR	09-NOV-1999; 99US-0164258.



XX PA (GLAX ) GLAXO GROUP LTD.  
 XX PI Kimmerly WJ;  
 XX PF WPI: 2001-316495/33.  
 XX PR N-PSDB; AAH52873.  
 XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 XX PT useful for vaccinating against infections, e.g. endocarditis -  
 XX PS Claim 18; Page 330; 2188pp; English.  
 XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX SQ Sequence 303 AA;  
 Query Match 6.9%; Score 80; DB 22; Length 303;  
 Best Local Similarity 24.6%; Pred. No. 2.1;  
 Matches 45; Conservative 22; Mismatches 48; Indels 68; Gaps 12;  
 QY 67 PWINGQKKPLY-----LYGAFLMNPALAKATKTTLNKGNL-----AWFIG 106  
 Db 141 pwvieesktvfesrpllldnnhrylygflh-----qingkeilmedlws- 191  
 QY 107 GTLGLRKRAGDSATRVVEALSVPEDVSGIGRG-----NLLKF-----WFAQ 152  
 Db 192 -----lesmndyeklyltylvqgltnklidf--ihrgmqrylnfkfkkyntslftdwnq 244  
 QY 153 A---IAANYDPKEANSFTNYKGFSAFYMYGITSLSFRAYGAYSKPANDK-LGSDFTFRK 208  
 Db 245 aemiaaenvdlvdrv-----aafvy-----lsyr-----rsgqptkrlqmdldfnvrsr 290  
 QY 209 FDL 211  
 Db 291 ykl 293  
 RESULT 10  
 AAG82824  
 ID AAG82824 standard; Protein; 359 AA.  
 XX AC AAG82824;  
 XX DT 03-SEP-2001 (first entry)  
 XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:2742.  
 XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
 XX KW vaccination; endocarditis.  
 XX OS Staphylococcus epidermidis.  
 XX

PN WO200134809-A2.  
 XX PD 17-MAY-2001.  
 XX PF 09-NOV-2000; 2000WO-US30782.  
 XX PR 09-NOV-1999; 99US-0164258.  
 XX PA (GLAX ) GLAXO GROUP LTD.  
 XX PI Kimmerly WJ;  
 XX DR WPI: 2001-316495/33.  
 XX DR N-PSDB; AAH53674.  
 XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 XX PT useful for vaccinating against infections, e.g. endocarditis -  
 XX PS Claim 18; Page 714; 2188pp; English.  
 XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX SQ Sequence 359 AA;  
 Query Match 6.9%; Score 80; DB 22; Length 359;  
 Best Local Similarity 24.6%; Pred. No. 2.7;  
 Matches 45; Conservative 22; Mismatches 48; Indels 68; Gaps 12;  
 QY 67 PWINGQKKPLY-----LYGAFLMNPALAKATKTTLNKGNL-----AWFIG 106  
 Db 197 pwvieesktvfesrpllldnnhrylygflh-----qingkeilmedlws- 247  
 QY 107 GTLGLRKRAGDSATRVVEALSVPEDVSGIGRG-----NLLKF-----WFAQ 152  
 Db 248 -----lesmndyeklyltylvqgltnklidf--ihrgmqrylnfkfkkyntslftdwnq 300  
 QY 153 A---IAANYDPKEANSFTNYKGFSAFYMYGITSLSFRAYGAYSKPANDK-LGSDFTFRK 208  
 Db 301 aemiaaenvdlvdrv-----aafvy-----lsyr-----rsgqptkrlqmdldfnvrsr 346  
 QY 209 FDL 211  
 Db 347 ykl 349  
 RESULT 11  
 AAY17217  
 ID AAY17217 standard; Protein; 327 AA.  
 XX AC AAY17217;  
 XX DT 03-AUG-1999 (first entry)  
 XX



```

CDS
1892..2725
/gene="rs2"
/codon_start=1
/transl_table=11
/product="S2 Ribosomal Protein"
/protein_id="AADI8835.1"
/db_xref="GI:4376999"
/transl_table="MESOSCKLTIKDLMSAGAHGHTRRNPKMKLYIFEKNGLYI
INLAKTQOLNALPHIRKVIQDNKTVLFVCTKKAQCVIREAAIEAGEFEFIAERWLG
GMLNTMIRNSIKTLDKIEKDLRSNAYLTKKAALLAKRHQLRNLNLESGIRTWKKA
PGLLVDPSTPEKTAIAEAKKLGIPVLALVDTNCDPTDHIIPCNDDSLKSLRIIN
VIKENIEIAKHKLGIETVSPVKSLEVPDLSAFESSQDDSEENREEDLLAKKFDGEA
N"
2725..3573
/gene="tsf"
/notes="Cp0697"
2725..3573
/gene="tsf"
/codon_start=1
/transl_table=11
/product="Elongation Factor TS"
/protein_id="AADI8836.1"
/db_xref="GI:4377000"
/transl_table="MSDFSMETLTKLRQOTGVGLTKCKEALACGNGLEEAIVYLRLK
GLSAGKHEHRETEGIIIAKTNDANGTALIEVNVETDFVANNVFRFVSLLNDILK
YKVDTVLEALSOAASQPSLSVDELAVTMOTVGENIRISRVAFPRATNSTVTGISH
NGKTVALTMLSGSSTADSLAKDIAMHVAAQPOFLSKESVPAEAIKAEKEVIAIOIQ
GKPOREIEKIYTGKLTFFQEAELLEQPFINKADLSIQSLDDFSKTSGSSVAIEQFI
LWKIGA"
3581..4327
/gene="pyrH"
/notes="Cp0698"
3581..4327
/gene="pyrH"
/codon_start=1
/transl_table=11
/product="UMP Kinase"
/protein_id="AADI8837.1"
/db_xref="GI:4377001"
/transl_table="MAKOTRRVLKISGEALS KDSSNIDEMRLSLVSELRAVRNND
IEIALVIGGNLRLCLAEQKELQINRVSDOMGLATLINGMAVADALKAEDICLLT
STLSCPQALDYLTQKSEALDQKILCTTGAGSPYLTDTTGALRACELNLDVLIK
AIMHDEGVYDKDPLFPDAVYDFVSKDFLSNLQGVMDASAILCMLDMSHIPRVFSP
LOHSLKALFDPTTGTLVSEDVNVHVCSPRH"
4305..4847
/gene="trf"
/notes="Cp0699"
4305..4847
/gene="trf"
/codon_start=1
/transl_table=11
/product="Ribosome Releasing Factor"
/protein_id="AADI8838.1"
/db_xref="GI:4377002"
/transl_table="MSVLQDTEKMAAALDFHKEVKSFRFGKHAHPALVIVVDVYG
TMRSLDSIASVADRLQVISPIDGNNSAIAKGIITANLNLPQEVGSLIRIKVPE
PTADYRQMIKOLRRKCEAKINVRNIRREANDKLKDSALTEDVYVGNKKEKIQELTD
KFCQQLDELTKQKEATASI"
complement(4945..5019)
/product="tRNA-Glu"
complement(5045..5117)
/product="tRNA-Lys"
5220..5735
/gene="Cp0700"
5220..5735
/gene="Cp0700"
/codon_start=1
/transl_table=11
/product="CT676 hypothetical protein"
/protein_id="AADI8839.1"
/db_xref="GI:4377003"
/transl_table="MVHSPTHQCYHCQOPATICYTEIDKDKVIRSYCATCPQPSHYH
NNEHLSLSKGVGLTLECGNCKTVHHSKQDDEQLLGCHQCYTNFKNQITSKLSERVV
SSSFTMEKGSGSLHGRAPGEASNTNPLLKIALNEALQDTEREDYEQAAVIRDOIN
HLTKNPDPS"
5722..6798
/gene="karg"
/notes="Cp0701"
5722..6798
/gene="karg"
/codon_start=1
/transl_table=11
/product="Arginine Kinase"
/protein_id="AADI8840.1"
/db_xref="GI:4377004"
/transl_table="MTLPNDLLETLVKRKESPOANKVMPVTTFSLARNLSVSKFLPCL
SKEOKLEIQTITSHFNHIEGFEFIVLPKDTPLMOKEFLLEHLEHLLPYDLVGNPEGE
ALVYRSRGDEFLAANFQODHLVHGIDFQGNVEKFTLDQLVQDLSVLSKLSFAFSSEFG
FLTNPNKSGTGLKRSQCFLLHIPALLIYSGFNTLDEVEIITSSLLGVGTGPGNIVV
LSNCSIGLTFEELLSRLITASKLSVAEAAKRLSEENSGDLKLLILSLGLLTHS
COLEKETLDALSWIOLGIDGLIKVTENHPLNPLFWQIRRAHLALQKQAEASRDLO
KDTISHLRASVLKELTKGLSPESF"
complement(6831..6903)
/product="tRNA-Thr"
complement(7049..9808)
/notes="Cp0702"
9808
/gene="yscC"
/codon_start=1
/transl_table=11
/product="probable Yop proteins translocation protein
C/general secretion pathway protein"
/protein_id="AADI8841.1"
/db_xref="GI:4377005"
/transl_table="MKTVILNIGRKILQGIKKKKKIGLSGLFELDLVLLGVSSORP
TETSANVKNLRLDEKLAACPKNSAASLSAKSHYKTTGPGSIPKSVSKFQDQKTF
QKTSGSAFPAPPTTLKELEKRRKPRERTADVKRSPRLPTQVEEVPVPAASKQEL
DSIQWEEKQNYARRVAINLSIKOLEQTSTVEKDVQPKTQATPHASKKNVASP
STNPGIEKAATTVAVPQDKSEEEKKERTKRELTCELDKNGYTVNFEDISLELL
QFVKSIGTNEVFDSDNQFNVTIVSHDPTSVDDLSLILQLVKHMLKVVEQGNVVL
IYRPHLSKLSVTVDSSLKCEAVVTVRVLVSVSPSAVNI IOPLLSHDAIVSA
SEATRHVILSDIAGNVKPSLLAALDCPCTSDMTEVEVYKYPANPAALVSCQVLTG
LAEDDAQMFQTPQNTKIFVYSSPRLANKAQLKSLDVPEDMTDLPDASTALGG
TGTPSPKSLRFMYKLKYNGEVIANALQDQIGLYIVTAMDEFTINTLNSIQWLEVN
NSIYIIGNQNVNVIIGLLGLDLPKQVYEVILDTLSLEKSWDFQWVALGDEQS
KVAYIIGNNVTGATPTKATVPPTPGPSIPLTPGQLTGFSDMLNSSFAGLGII
GNVLSHGKSFLLTGLLSALDQDQDVIIVLNPRIMAQDTQOASFVQVPTVDTKT
IIQFTGTVONIDYIGVNLVVTSTVAPNPNVTLOIQTISELSASGSLTPVDTKT
YAATFLOIPDCGFLVMSGHIRDKTKVYGVPLNLSIPLINGLFSRTIDQRQKNIMM
FIKPKVISSPEEGTRVTNKEGIRYNWEADEGSMQVAPRAPECOGPPSLQAESDFKII
EIEAQ"
complement(9805..11313)
/gene="Cp0703"
complement(9805..11313)
/gene="Cp0703"
/codon_start=1
/transl_table=11
/product="S/T Protein Kinase"
/protein_id="AADI8842.1"
/db_xref="GI:4377006"
/transl_table="MDCRGIPPEPOVIGGYHVKKILSKLRSRVVHGLHPETRHS
VIKVESPSSETSRSVYNFLKEAQLSHQIHPNIVKPHRYGKWQDCLYIAMEYIEGIS
LREYILAQFISLPQADIIIFDIALEHLNRNHLKDKIPKPNILITPQGIKILDFG
LADWDEIQRAPSVIGTPYVSPQSGSHSPASDIYALGGLAYELILGHLSLGRV
FLSLVPERISKILAKQPSNNRYSSTREFIQDHHYRMSGDMQEDLRKDHDTVALY
EQLOTQRFWLPALTRFPDFLSGVLYHGYPLYPHAYDTLLEGVDFNVLWLSYISNA
TIALSVYKSLVCCQDLOPLDRVCEINCEILRMKIPIDEMGI SILCLEISKENKELS
WIAGKTVFWIKRQGRVVDPESEFSGLGKITSQIRETKVAMEIGEVNCTLELEE
SVASLKLTLSEALQDRQKALFCPIESIHGGIQSGSGNSPSTLISUKRIR"
complement(11329..12444)
/gene="fln"
/notes="Cp0704"
complement(11329..12444)
/gene="fln"
/codon_start=1

```



```

/transl_table=11
/product=Flagellar Motor Switch Domain/YscQ family"
/protein_id="A018843.1"
/db_xref="GI:4377007"
/translation="MAVAADSSASWLNKRNFLSSIGKTEQVAAPFPEKELCQHKIR
EKPRLEDVQVSIKFRGSITAVEATKEFVHLLIQPMVQFWEVENLLFTSEDLQEL

```

Query Match	99.78;	Score 1098;	DB 1;	Length 20057;	
Best Local Similarity	100.0%;	Pred. NO. 1.2e-275;			
Matches 1098;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	4	ttgctgtagggaacccctctgaccagcttatttaattggtgtacaaatggaaggt	63		
DB	408	TTGCTGTAGGGAACCCCTCTGATCCAGCTTATTAATGATGTTACAAATATGGAAGGT	467		
QY	64	gtcgagagatccttgagactcttgctgacttggtgctgagcgtctattagcttgcgt	123		
DB	468	GCTGCAGGAGATCCTTGGGATCCTTGGCTACTTGGTGGACGCTATTAGCTACGCTGCT	527		
QY	124	gattttacgagactatgtttgacacctatcttaaaagttagatgacacctaaacattt	183		
DB	528	GGATTTTACGGAGACTGTGTTTCACCGTATCTTAAAGTAGATGACCTTAAACATTT	587		
QY	184	tctatggagaccaactctactgactgctgctgcacaaactactactgctgctgtagata	243		
DB	588	TCTATGGGACCAAGCCTACTGGATCCGCTGCTGCAAACTACTACTGCGGTAGATAGA	647		
QY	244	cttaaccggcctacaataagaattacacagatgagagtggttcacttaagcagcttc	303		
DB	648	CCTAACCCGGCCTACAAATGAAGCATTTACAGATGACAGTGGTGGTCAATGCAAGCTTC	707		
QY	304	attgcttaaacatttgatcgcttgaattgtttcttactttaggagcttctaagtt	363		
DB	708	ATTGCTTAAACATTGGGATCGCTTGTATGTTTCTGCTACTTTAGGAGCTTCAATGGT	767		
QY	364	tacattaggaatactcacagcttcaactctggtttatttattcggtggttaaggtact	423		
DB	768	TACATTAGGAACTCTACAGCTTCAATCTCGTTGGTTTATTCGGAGTTAAAGGTACT	827		
QY	424	actgtaaatgcaaatgaactacaaacgctttcttaagtaacggagtggtgaaacttac	483		
DB	828	ACTGTAATGCAATGAATACACAACTTCTTTAAAGTAACGGAGTGTGTAACCTTAC	887		
QY	484	acagacactcttctctggagcgtgagcgtgctgagcgttattggaatgcggttgt	543		
DB	888	ACAGACACTCTTCTCTTGGAGCGTAGCGCTGCTGGAGCCTTATGGAATGCGGTGT	947		
QY	544	gcaactttggagctgaattccaatatgcacagtcacaaacctaaagttgaagaactta	603		
DB	948	GCAACTTTGGGAGCTCAATTCGAATTCACAGTCCAAACCTTAAAGTTGAAGAAGCTTA	1007		
QY	604	gtgatctgaactatcgcaattctctgtaacaaacccaagggctataaagcgctgt	663		
DB	1008	GTGATCTGAACATATCGCAATCTCTGTAACAAACCAAGGGCTATAAAGCGGTGCT	1067		
QY	664	ttcccttgcacacagcgtgctgtagcaacagctactggaacaaagctcgcgaccatc	723		
DB	1068	TTCCCTTGCACACACACGCTGCGTAGCAACAGCTACTGGAACAAAGTCTGGACCATC	1127		
QY	724	aattataatgaatggcaagtaggagcctctctattctacagactaaactcttagtgcca	783		
DB	1128	AATTATCATGAATGGCAAGTAGGAGCCTCTCTATCTTACAGACTTAACTCTTTAGTGCCA	1187		
QY	784	tacattggaatacattggtctgacgaacttttgatgctgataacacccgcatctgctcag	843		
DB	1188	TACATTGGAGTCAATGGCTCTCAGCAACTTTTGTATGCTGATCAATACATCCGCATGCTCAG	1247		
QY	844	ccaaactactacagctgttttaaaactaaactgcatggaacacctcttcttactaggaat	903		
DB	1248	CCAAACTACTACAGCTGTTTTAACTTAACTGATGCAATGCAACCTCTTTTACTAGGAAT	1307		
QY	904	gcaacagactgttactactgattcgtttctcagaacttcatgcaaatgtttctcctcag	963		

DB	1308	GCACAGCATGTCTACTACTGATTCTCTCAGAGCTTCATGCAAAATTTTCTCTGTCTAG	1367		
QY	964	atacaaaagttaaatctgaaagactgtgaggtactgtgagagctactttagttgat	1023		
DB	1368	ATCAACAAGTTTAAATCTAGAAAGCTTGTGAGTTACTGTAGGACTACTTATTGAT	1427		
QY	1024	gctgataaatggctcactactcagagctcgcttatttaataacagagagctgctcagca	1083		
DB	1428	GCTGATAAATGGTCACTTACTGCAAGAGCTCGTTTAAATTAACGAGAGCTGCTACGTA	1487		
QY	1084	tctggctcagttcagattc 1101			
DB	1488	TCTGGTCAGTTCAGATTC 1505			
RESULT	5				
AE002167/c					
LOCUS	AE002167	25150 bp	DNA	BCT	06-APR-2001
DEFINITION	Chlamydomophila pneumoniae AR39, section 4 of 94 of the complete genome.				
ACCESSION	AE002167	AE002168	AE002161		
VERSION	AE002167.2	GI:8163349			
KEYWORDS	Chlamydomophila pneumoniae AR39.				
SOURCE	Chlamydomophila pneumoniae AR39.				
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.				
REFERENCE	1 (bases 1 to 25150)				
AUTHORS	Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.				
TITLE	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39				
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)				
MEDLINE	20150255				
PUBMED	10684935				
REFERENCE	2 (bases 1 to 25150)				
AUTHORS	Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712				
COMMENT	Medical Center Dr, Rockville, MD 20850, USA				
	On Jun 1, 2000 this sequence version replaced gi:7188982				
FEATURES	gi:7188971.				
source	Location/Qualifiers				
	1..25150				
	/organism="Chlamydomophila pneumoniae AR39"				
	/strain="AR39"				
	/db_xref="taxon:115711"				
	/note="synonym: Chlamydia pneumoniae AR39"				
gene	306..1325				
	/gene="CP0032"				
CDS	306..1325				
	/gene="CP0032"				
	/note="similar to GB:M25323 SP:P13580 GB:M30785 GB:X17434				
	PID:146330; identified by sequence similarity; putative"				
	/codon_start=1				
	/product="glutamyl-tRNA reductase, putative"				
	/protein_id="AAF37927.1"				
	/db_xref="GI:7188972"				
	/transl_table=11				
	1692..2084				
gene					
	PIPLTCHRAELVYVSEPRTAQAALISLTSGIRYRHRGLSCFHLQVTSIGD				
	LIFGETELQGVKRAYLKGSKERELPDFLHFLKQKALKEGVEYRFPDQVTV				
	VVOEILLSDKSIYTNFLVGYSDINKRVAIYLFGHHRITFCSROQVATPYTL				
	ETLSFRQPYDVFIFGSEASQFSDLSCLASIPKRVDFDNVPRFLMKETGTV				
	YLIDIFISECVQKRLOCTKEGVNKAALLLTCAARKOWEIVEKSSHTQTQOISS				
	SVLSY"				



```

/ gene="CP0033"
1692..2084
/ gene="CP0033"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37928.1"
/db_xref="GI:7188973"
/translation="MLEKLKNFATYMGITSTLELDAGAVLPISEVVKVRAQONAD
NEVLISLALGPSADTAKLYLQMMIGNLFGRETGGALGLDSEGNVVMVRRSGDT
TYDDFVRHVESFMNFSETWLSDGLGKQ"
2098..4635
/ gene="CP0034"
2098..4635
/ gene="CP0034"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37929.1"
/db_xref="GI:7188974"
/translation="MAVRLIVDEGLSGVIFVLEDGISWISGRSSANDIPIEDPKLG
ASQALINTDGSYITNLDITPIVNVGVAIQETQLKNEDTILGNSQYSLDFEFD
PQDLVYDIPENFSDNSDGSNQGKLEPROTSETNHSKPKKLLTKDOGSDD
PITSGDQLADAFASAKAEKNOPRAKVAKKGLKESSEINPKPEONAKDSPKEERT
NKQNAIMEDGASPRODPKSAEPSLNTARDETPLENKPVEEKANKKATPDSPE
KKDQPEESKEGSKIETPLDSOKESDEKAEAFVQEEENITENKEDSSAADA
NDTASDTHADNKETPKKVENKSAVLSVPHVQDLFRPDQTFIPAEIDDIJAKKNISV
DLTQSRFLLVLAGANIGAEFHLDGKTYILGTDPTCDIVFNDLSVSHQAKITVG
NDGGLIEDLSKNGVIVEGRKIDKSTLSSNVVALGTTLLFDHHPADTIVASL
SPDYSLFGRQQAALREQAQEEKOKRATLPAGSFILTLFVGGAILFGIGTAS
LFTKEVPLENIDYODLAOVINQFPVRYTENKTSOLFGLGHVKNSTKCYSELYK
VDHFSKVSVDNDVIDEAVQEMNILLSKPEEKGISHSGPEKFLITCYVITEEQ
RACLVDYLNIHFNVLSENKVVYETQMLKALHGLLOGGFANHVAIVNGEVLITGY
VNDDAEAFRAVVELSGIPGVLNVKFAVILLPAEEGILDLNLRYPNRYRTVGSRYG
EISINVNVGRLRGDVIDGMTVTSIQPNAIFLEKEGLKYKIDYNK"
4682..4930
/ gene="CP0035"
4682..4930
/ gene="CP0035"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37930.1"
/db_xref="GI:7188975"
/translation="MFNMNTAKEKNQSPLLDLEQDMQDHRAQELKASVQDKVHKL
HALLRGSDKESFGQQQLLAGYVALQVLRINKKMI"
4952..5206
/ gene="CP0036"
4952..5206
/ gene="CP0036"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37931.1"
/db_xref="GI:7188976"
/translation="MATNKSCTAFDNKMLDGVCTYVKGVOYLTELETSTQGTVDLG
TMFNLFQRMQLLSQYMESVSNILTAVNTMITMARVKG"
5225..5674
/ gene="CP0037"
5225..5674
/ gene="CP0037"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11

```

```

/product="conserved hypothetical protein"
/protein_id="AAF37932.1"
/db_xref="GI:7188977"
/translation="MADLEVFQADFALLFEAGLIAIKQGDSDSARKLQPSUHLINPNH
YGHDLGALISLHKMDLFAEERLSALIKGNEDNMSKAFSLTHMLTLVHQSSFEV
RRSELSCLKFAFDQVIANCKIESTRALAQSVLDNHDTLVAKSAGPLG"
5708..6382
/ gene="CP0038"
5708..6382
/ gene="CP0038"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37933.1"
/db_xref="GI:7188978"
/translation="MIDPEVCFPNLDGAEASQTSQNSGTPLASELKKDISPFALGSY
AAPKDTTLVQGFKNPMAMQDQNSNLDLPELQALESELEQEQINNKGRLWDFRST
FEDSOTTAOFADHFQAVGVIIDILNEDLNTIAEHTQODARKEGKSVTRKIIDWV
SSGEVLNRALLYFSDRDNRESLANFLKVQYAVQRAELFASIVGTSVSSVKTI
MTTQLG"
6384..7712
/ gene="CP0039"
6384..7712
/ gene="CP0039"
/ note="similar to SP:P21211 GB-U02499 SP:P21212 SP:P40290
PID:437202; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="type III secretion cytoplasmic ATPase Scn"
/protein_id="AAF37934.1"
/db_xref="GI:7188979"
/translation="MDQLTDFVLSQGLDNLTVVGRITVEVGMILKAVPNVRV
GEVCLVKGNEPLVTEVVGTSQFAELSPGLSGVSPSEVPTGCLPIHNRAGNL
LGRVLNGEPIVETKGPLQNDVDTFFIRAPDPLHRAKLQILSTGVRICIDMLT
VARQINGIFAGAGVGKSSLLGMIAARNAEADVNIALIGREGRVREFIEGDLGEG
MKR5VYVSTDSQSLNAAAYTAIAYFRQGTVTVLMDSVTRFARALREVGL
AAGVPPRAGYTPSVFLLRSGASDKGTITAFYTVLVAGDMMPEPVADEVKSI
LDGHVLSNALAQHYPAIDVLASISRLTAIYPEQORRIIGKAREVLAKYKANEML
LRIGYRRGSDREIDFADHIDKLNRLKQDIEKNTYEEAAQQLAIRF"
7730..8236
/ gene="CP0040"
7730..8236
/ gene="CP0040"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37935.1"
/db_xref="GI:7188980"
/translation="MAKYPLEPVLAKKRDVRAEKVVVKEKRRLLIEBQELREKAE
RDVKNHYMQIKOQLRDLDEGTSADVLOIKSVIKVAVOLSEEEKVNRKQEVLA
ASKELEKAEVNLAKRKEEETRLHKEEMKEALKEAREEKEQDEMGQLLFLQRK
KKRESGGS"
8240..9082
/ gene="CP0041"
8240..9082
/ gene="CP0041"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37936.1"
/db_xref="GI:7188981"
/translation="WELAKTAEISAKTDNHTVYQNSPEPRDSRDVVFSLGKQTR
QEKTSKGNTRTESRPFADEEKRVDDIEAVGSEKEQSEOEAEENAFAGMSLID
IAAGSAEAVVEVAPIAVSSITQMIENIILTSVESMVISINQLVELVLDASSV
PEAFVGANLTLVQSQDLVSRKFSFVDATMAEADLVTNNPQLSSLSVLSALKGHOLT"
99.7% Score 1098; DB 1; Length 25150;
Query Match

```



Best Local Similarity 100.0%; Pred. No. 1.2e-275;  
Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttgccttagggaacccctctgatcccaagccttattatgtatgtgtacataatggaaggt 63  
|||||  
Db 21128 TTGCCTGTAGGGAACCCCTCTGATCCAAAGCTATTAATATGATGTTGATCAATATGGAAGT 21069

QY 64 gtcgaggagatcccttgatcccttgctgctactgtgtgcagcgtatttagcttgcgt 123  
|||||  
Db 21068 GCTGAGGAGATCCCTTGGGATCCCTTGCCTACTTGGTGGCAGCGCTATTAGCTTACGTGCT 21009

QY 124 gqattttacggagactatgttttgcacgctattcttaaaagtagatgcacccctaaacatt 183  
|||||  
Db 21008 GGAATTTACGGAGACTATGTTTTCACCGCTATCTTAAAGTAGATGCACCTAAACATTT 20949

QY 184 tctatgggagccaaagcctactggatccgctgctgcaactactactactgcgctagataga 243  
|||||  
Db 20948 TCTATGGGAGCCAAAGCTACTGGATCCGCTGCTGCAAACTATACCTACTGCGGTAGATAGA 20889

QY 244 cctaaccggcctacaataagcatttacacatgcagatggttcaactaacgcaggttc 303  
|||||  
Db 20888 CCTAACCCGGCCTACAATAAGCATTTACAGATGCAGATGCTTCACTAATGCAGGCTTC 20829

QY 304 attgccttaaacatttgggacgtcttggatgttttctgtacttttagggagcttctaagt 363  
|||||  
Db 20828 ATTGCCTTAACATTTGGGATCGCTTTGATGTTTTCTGTACTTTAGGAGCTTCTAATGCT 20769

QY 364 tacattagggaaactctacagcgttcactctctgttatttcggaggttaaggtact 423  
|||||  
Db 20768 TACATTTAGGAAAGCTCTACAGGCTTCAATCTGTTGTTTATTCGGAGTTAAAGCTACT 20709

QY 424 actgtaaatgcaaatgcaactacccaacgcttcttaagttaacgaggtgttgaacttac 483  
|||||  
Db 20708 ACTGTAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 20649

QY 484 acagacacctcttctcttgagcgttagcgtctgagcgtcttggagccttatgggaatgcggtgt 543  
|||||  
Db 20648 ACAGACACCTCTTCTCTTGGAGCTGAGCGCTGCTGAGCGCTTATGGGAATGCGGTTGT 20589

QY 544 gcaactttggagcgtgaattccaaatgacacagtcacaaacccctaaagtgaagaactaat 603  
|||||  
Db 20588 GCAACTTTGGGAGCTGAATTTCCAATATGCACAGTCAACACCTTAAAGTTGAAGAATTAAT 20529

QY 604 gtgatctgaacgtatcgcaattctctgttaaacacaccccaagcgtataaagcggtgt 663  
|||||  
Db 20528 GTGATCTGAATATCGCAATCTCTGTAAACAAACCCCAAGGCTATAAAGCGCTGCT 20469

QY 664 ttcctcttgcacagacgctggcgttagcaacagctactgtggaacaaagctcgcgaacct 723  
|||||  
Db 20468 TTCCCTTGCACACAGACGCTGGCTGAGCAACAGCTACTTGGAAACAAAGCTCGACCAATC 20409

QY 724 aattatcatgaatggcaatgagagcctctctattctacagactaaactcttagtgca 783  
|||||  
Db 20408 AATTATCATGAATGGCAAGTAGAGCCTCTCTATCTTACAGACTAAACCTTTTAGTGCCA 20349

QY 784 tacattggagatacaattgctcagcaacatttttgatgctgataaactcgcattcgcag 843  
|||||  
Db 20348 TACATTGGAGTACAATGGCTTCGAGCAACTTTTGATGCTGATAACATCCGCATTCGCTCAG 20289

QY 844 ccaaaactaacactgctgttttaacttaacttaactgaatgaacccctcttacttaggaat 903  
|||||  
Db 20288 CCAAAACTACTACAGCTGTTTTAACTTAACTGATGCAAGCCCTCTTTTACTAGGAAT 20229

QY 904 gccacagcattgctactactactgcttctcagactctcactcaaatgttctcctgcaag 963  
|||||  
Db 20228 GCCACGATGCTGCTACTACTGATTCCTCTCAGACTTCAATCAATGTTTTCCTGCTCAG 20169

QY 964 atcaacaagttaaatctagaaaagcgttgaggttactgtgagagcacttagttgat 1023  
|||||  
Db 20168 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT 20109

QY 1024 gctgataaatggtcaacttactgagaaagcgtgtttaaataacgagagagcgtcctcaagta 1083  
|||||

Db 20108 GCTGATAATAGTCTACTTACTGCAGAACTCGTTTAAATTAACGAGAGAGCTGCTACGTA 20049

QY 1084 tctggtcagttcagattc 1101  
|||||

Db 20048 TCTGCTCAGTTTCAAGATTC 20031

RESULT 6

AP002547 300550 bp DNA BCT 08-DEC-2000

LOCUS Chlamydomydia pneumoniae J138 genomic DNA, complete sequence, section 3/4.

ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008

VERSION AP002547.2 GI:10176693

KEYWORDS

SOURCE Chlamydomydia pneumoniae J138 (strain:J138) DNA.

ORGANISM Chlamydomydia pneumoniae J138.

REFERENCE 1 (sites) Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.

AUTHORS Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.

TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA

JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)

MEDLINE 20330349

REFERENCE 2 (bases 1 to 300550)

AUTHORS Shirai, M.

TITLE Direct Submission

JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases. Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail: mshirai@po.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227, Fax: 81-836-22-2415)

COMMENT On Sep 15, 2000 this sequence version replaced gi:6172298 gi:6172300 gi:6172396 gi:6172398 gi:8978889.

AB033786-AB033787, AB033816-AB033817: Submitted (25-Oct-1999). Location/Qualifiers

FEATURES

source

1..300550

/organism="Chlamydomydia pneumoniae J138"

/strain="J138"

/db\_xref="taxon:138677"

/note="synonym:Chlamydia pneumoniae (strain J138)"

gene

complement(21..794)

/gene="dapF"

CDS

complement(21..794)

/gene="dapF"

/codon\_start=1

/trans\_table=11

/product="glutaminolactate epimerase"

/protein\_id="BAA98725.1"

/db\_xref="GI:8978890"

/translation="MAYSPSTISKYFIYSGAGNRFLIGETLPEVDVRFQCFRVD GFYLPKSSCADAOLITFNSDGRPTWGNLCRAIAHLASOKKSDISYSTDSGLYS GFYSPDRVLVDTLADWRASVHRLSPDPLKEVCIHTGVPHAVVILPEISTLDL SILGFLRYHTFSPDGNVNVNFVQLGHCQLRVRYTERGVEGETACGTCALASALVV SNSYWKESIQHTWGGELMTVVSQNRGRVYLQGSVTHDL"

gene

complement(763..1338)

/gene="clpP\_1"

CDS

complement(763..1338)

/gene="clpP\_1"

/codon\_start=1

/trans\_table=11

/product="CLP protease"

/protein\_id="BAA98726.1"

/db\_xref="GI:8978891"

/translation="MADGEVHKLPDIIKELLEARRVFPFSEPSASDAIKKLVIL ELKDPGRPIVFVINSVPGSVDAAGFAVNDQIKMLSPVTVTGLAASNGSVLSCLAP GRRTATPHSRIMIHOPISIGGPITGQATDLDIHAREILKTRIIDVYVEATNQPRDII EKAIDRWMTANEAKDFGLDGLFNFNDL"

gene

complement(1358..2917)

/gene="glyA"

CDS

complement(1358..2917)

/gene="glyA"



```

/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/transl_table="MLKVFEKFKFAIVEIFTKVAVVSVLLHKLFLKLENAAGKGGKQSLAS
TAYLAADHLNAPFSGIGERIIDELKSORSHLKMIASENYSSLSVOLAMNLLTDKVC
EGSPFKRYSCCENVDALIEWECVETAKELFAACACVOPHSGADANLLAVNAILTHKY
QGPVAKGYKTYNLTETEEYTLKKAENSSCVCLGPSLNSGCHLTHGNVRNUNWSKLM
RCFPVDNPTCEDYDAISRLAKEIKPKVLIAGTSSYRRRLNFAVLKQIAEDCGSVL
WYDMAFAGLVAGGVFVDEENPIYADIIVTTTHKLRGPRGGLVLAATREYESTLTKA
CPLMGSPFHVIAAKTVALKEALSVDPKYAHQVNNARLARERLSHGLRLITGGT
DNHMMVLDLGSISGKIAEDILSSVGIANRNSLPDAIGKWTSGRLGTPALITL
GMGIDEMEVEDIIVKVLNRNRLSCHVEGSSKKNKGELPEAIAQEARDRVRNLLRFP
LYPTDLEALV"
3107..3778
/gene="CPj0522"
3107..3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/transl_table="MTLYLGNKQTKARKYQAHYLPILTLFPAKSTPQNKRALQFLPQ
ATHVLTSPSTHLFLSRMTSLLSKATLTKTYLCLIGESTKERILSLFLGQVYVVAQ
ETAEIGFPLQALPSSARILYPHSSLARPVIREFLYNRFTFSYPHYTVKPKLKNKI
LSKYKKIIFTSPTSVRAFAKIFPREPKTYWCQGMTLQEFQKFSQKQVSLLETGK
SRTSP"
3843..4175
/gene="CPj0523"
3843..4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/transl_table="MASSATPGFGDTAPSLPFPATPRPNFKALFVTAIALVNIAL
IATTAIGLCIHPLCSFIFLPAIFLYFISRYICSHYARNVYIALDVVDPHSLQDMRS
HSP1FSDR"
4223..5302
/gene="CPj0524"
4223..5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/transl_table="MSGPSRTSESSQVSVLSPVRDKETAPKPKQFTIAKISTLAILASL
ALGALVAGISLTVILGNPVFLALITLALFSVTVFLVYHOMTSKVSSNQKVLQNKPK
PLGAWQEKVDCYSNEMQFVNNHLPKFAVIAOTDASQFPQFTLGLRYIEKNQST
GILNPVCGPTNLIDNTANLSTILYSLTKDKSVMDTKQREGGPAKGDPP$PEVRV
VKLPNEALDQFNLNLSAERKSILPTFLHVCQKSEELPNQOYIYQOALLAYENCL
KAALIESHAALVALPLFTSVIEVPPEILPKEGTIFYDNTQOAFCKRALLDAIQNTALR
YPOKSLVILADPNTIESQSRSE"
5642..6406
/gene="CPj0525"
5642..6406
/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="CT398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/transl_table="MHDALLSILAQELDKMIRLMRVKKEHOKELAKVQSKSDIRR
KVQKELEMENLKTQIRDGENRIQISEQINKLENQQAARKMDFFNALTQEMTANK
RRSLHOLSDMLKQAGGEDLYSLKESLASTENSSSVIEKEIFESTKINEEGKAL
LEQRLKHAHNPELLSIYERLLNKKNDRVVPTNVCSCGCHVLTTPQHENLVKRD
RLIFCEHCSRILYQESQVNAQENSTAKRRRRRAV"
complement(5827..7816)
/gene="kpsf"

```

```

CDS
complement(5827..7816)
/gene="kpsf"
/codon_start=1
/transl_table=11
/product="GutQ/KpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/transl_table="MPSPMISTDVQCQDILGKQKAVDFFFOAFQPKAMQALAEKILGH
SGWFFSGVSGCVARKLVATLQSLSERALFFSPDLLHGDGLGLVSPGDIIVCLFSSK
GETOLDTVDPHLKSRAILVAITSMYPNLAALSDLVILPVAELDPFNLPTNST
TCOMIFDGLFAMLEFHSRGVSLSTYGNHPSGQVGMKANGKVDKMFPTKTEVPFCHLG
TCQIFDGLFAMLEFHSRGVSLSTYGNHPSGQVGMKANGKVDKMFPTKTEVPFCHLG
NPKCITEDSDIAIALQMLMESSPPAVLPVLDNEENRHVTGLLHMHTLAKACLL"
complement(7843..9024)
/gene="sucB_2"
3107..3778
/gene="CPj0522"
3107..3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="dihydrolipoamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/transl_table="MIFEFPPKIGETSSGGSIVRWLKNLGDHVARDEPLIEVSTDKI
ATELSPKAGRLRVFCVNEGDEVDGLGLEIEEISEADESTSCPLTSCETKSEA
GSSSSVMFSPAVLSLAQREGIDLNQKIAGTGGGRTVQDLEAVISEQQVSIPE
IFQCEVNRIPMSPLRRATIASLSKSSDEVHSLVVDVDTDLMLNLGSGRQRFDLTH
GKVLITTSFVQCLAOTLRQFPLNGSLDGTIVMKSVNVGVAVNLNKEGVVVPVTH
NCQDGLVSIKALADLSSRARLNKLDESEVQDSVTVNFGMTGALIGMFIIRYPEV
NAILGITIQKRVVVRDDDSLAIKRMVYVTLTFDHRVLDGIYSEBFLTSLKNRLESVTM
G"
complement(9038..10282)
/gene="gltr"
complement(9038..10282)
/gene="gltr"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/transl_table="MKLMKIFIGLVGVTLGLVLEDKAIEFKPIGIDIFLNLSSVYV
PLVCSWVIGTASISDKMKKIGIGISVGLYGLTITLTAALVIAIGLFAIFPSGNGCDFA
QAQMSDAVTVIDSNKTAAYFLSLIAQVPSNPVSAEAGNIIQIIIFAILGIALRL
SGERPVPERIDGSEIMLRVMNINMSFAPYGVGASNAWISGHNGLVWLQGLKFI
AYLACLFLHATLVFGGLVFCGKMSFSKFLSSMDAISCAVSTASSATPLVTRCVS
KNLGSVAESGVFLPLGATVNMGTALFOGAAVFAQAYNCPLSLSLLLVVTATF
SAGSAGVPGGGMITLGSVLASVGLPIQGIALLAGIDRLDIVGTPMNLGDVAVVATY
VASGEGLSPYESIKQESVET"
complement(10282..11379)
/gene="ycha"
complement(10282..11379)
/gene="ycha"
/codon_start=1
/transl_table=11
/product="Atpase"
/protein_id="BAA98735.1"
/db_xref="GI:8978900"
/transl_table="MKRFPSTFLFYRVRTIAISLEGILCWGLSGLLSKVFAFLVA
CWNRFSTPYRARTSVISGVNIVVGAGKPTVILWLAELRLRGYCGVLSRGYKSO
SSROKLTIVDSKVHSASVGDPELLMAELKPEGSVWVHKDRRIISAARAEKFGILL
DDGLQYRLKHDEIVAVVNGODPLGRRAFFPKRLRDFPLRLKTVDAIIVNGGKEAG
TVKRVSNAPQIFVKYTVIASVWTHGERIPKEALRELVRGVFCGLGPGFQGLNMLRE
EGHILGKLYLPDHAATIKELNITFCQOMANROGGLCTEKDYSKVLPRLSGEVSLP
IAKVMRLSVNQDDTSLNLNIEQIHKNRGN"
Query Match 99.7%; Score 1098; DB 1; Length 300550;
Best Local Similarity 100.0%; Pred. No. 1.3e-275;
Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 ttgcctgtaggaacccctctctgatacgaagcttataatgattggtacataatgggaagt 63
|||||
DB 179406 TTGCGCTGAGGGAACCCCTCTCATCAAGCTTATTAAATGATGGTACATATGGAAGT 179465
|||||
Qy 64 gctgcagagatccttgogactctgctgactgtggtgcgacgctattagctagctgct 123

```



|||||  
Db 179466 GCTGAGAGATCTTGGCATCTTGGCTACTTGGTGGCGAGCTATTAGTTACGTGCT 179525  
Qy 124 ggattttacggagactatgttttgcagcgtatctttaaagtagatgcacctaataacattt 183  
Db 179526 GGATTTTACGGAGACTATGTTTTCGACCGTATCTTAAAGTAGATGACCTTAAACATTT 179585  
Qy 184 tctatggagccagcctactggatccgctgctgcaaaactactactgcogtagataga 243  
Db 179586 TCTATGGAGCCAGCCCTACTGGATCCGCTGCTGCAAACTACTACTGCGGTAGATAGA 179645  
Qy 244 cctaaccggcctacaataaagcattacacgatgcagagtggttcactaaatgcaggcttc 303  
Db 179646 CCTAACCCGGCTACAAATAGCAATTTACAGATGCAGAGTGGTTCACCTAAATGCAGGCTTC 179705  
Qy 304 attgcttaaacatttgggagctgttggatgtttctgtacttttaggagctcttaaggt 363  
Db 179706 ATTGCTTTAAACATTTGGGATCGCTTTGATGTTTCTGCTACTTTAGGAGCTTCTAATGGT 179765  
Qy 364 tacattagagaaactctacagcgttcaatctcgttggttttatcgcaggttaaggtact 423  
Db 179766 TACATTAGAGAACTCTACAGCGTTCATCTCGTGTGTTTATTCGGAGTTAAAGGTACT 179825  
Qy 424 actgttaaatgcaaatgaaactacccaacgctttctttaaagtaacgaggtgtttgaactttac 483  
Db 179826 ACTGTAAATGCAATGAACACTACCAACGTTCTTTAAGTAACGGAGTGTGTAACCTTAC 179885  
Qy 484 acagacaccttctcttggagcgttaggcgtctgagcgtctatggagcgttatggagcgttgc 543  
Db 179886 ACAGACACCTCTTCTTGGAGCGTAGCGCTGCTGGAGCGTTATGGGATGCGGTTGT 179945  
Qy 544 gcaactttggagcgtgaattcacaataatcacagctcacaactcaactaaagtgaagaactaat 603  
Db 179946 GCAACTTTGGGAGCTGAATTCCAATATGCACAGTCCCAACCTTAAAGTTTGAAGAACTTAA 180005  
Qy 604 gtgactgttaacgtatcgcaattctctgttaaaacacccaagggcgtataaaggcgttgc 663  
Db 180006 GTGATCTGTAACTGCAATCTCTGTAAACAAACCAAGGGCTATAAAGGCGCTTGT 180065  
Qy 664 tcccccttccaacagagcgtgctagcaacagctactggaacaaagtctgcgaccatc 723  
Db 180066 TTCCCTCTGCCAACAGAGCTGGCTAGCAACAGCTACTGGAACAAAGTCTGCGACCATC 180125  
Qy 724 aattatcatgaatggcaagtagagcctctctatcttacagactaaactcttttagtgcca 783  
Db 180126 AATTATCATGAATGGCAAGTAGGAGCTCTCTATCTTACAGACTAAACTCTTTAGTGCCA 180185  
Qy 784 tacattggagtagaattgtctcagcaacttttgcgtatgataacatcgcattgtcgaag 843  
Db 180186 TACATTGGAGTACAATGGTCTCGAGCAACTTTTGTGCTGTATTAACATCGGCATTGCTCAG 180245  
Qy 844 ccaaaactaccacagcgtttttaaacttaactgcatggaacccctctttactaggaat 903  
Db 180246 CCAAAACTACCTACAGCTGTTTAACTTAACTGATGATGGAACCCCTCTTTACTAGAAAT 180305  
Qy 904 gccacagcattgtctactactgattctgtctcagactctcatgcaaatgttttcctgtcag 963  
Db 180306 GCCACAGCATGCTCTACTACTGATTCGTTCTCAGACTTCATGCAAAATGTTTCTGTCAG 180365  
Qy 964 atcaacaagttaatactctagaaaagcctgtgaggttactgtgagagctacttagttgat 1023  
Db 180366 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT 180425  
Qy 1024 gctgataaattgtcactactcagagcgtctgttaattaaacagagagagcgtctcagta 1083  
Db 180426 GCTGATAAATGGTCACTTACTCGAGAAAGCTCGTTTAATTAACGAGAGAGCTGCTCAGTA 180485  
Qy 1084 tctggtcagttcagattc 1101  
Db 180486 TCTGGTCAGTTCAGATTTC 180503

RESULT 7

CHTMOMP  
LOCUS  
DEFINITION  
Chlamydia pneumoniae major outer membrane protein (MOMP) gene,  
complete cds.  
ACCESSION  
M64064 M34942 M64063  
VERSION  
M64064.1 GI:144534  
KEYWORDS  
major outer membrane protein.  
SOURCE  
Chlamydia pneumoniae (strain IOL-207) DNA.  
ORGANISM  
Chlamydia pneumoniae  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. phila.  
REFERENCE  
1. (bases 1 to 1170)  
Carter, M.W., Al-Mahdawi, S.A.H., Giles, I.G., Trehan, J.D.,  
Ward, M.E. and Clarke, I.N.  
Nucleotide sequence and taxonomic value of the major outer membrane  
protein gene of Chlamydia pneumoniae IOL-207  
J. Gen. Microbiol. 137, 465-475 (1991)  
91237311  
Location/Qualifiers  
1. .1170  
/organism="Chlamydia pneumoniae"  
/strain="IOL-207"  
/db\_xref="taxon:83558"  
1. .69  
/gene="MOMP"  
1. .1170  
/gene="MOMP"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AA23143.1"  
/db\_xref="GI:144535"  
/translation="MKLLKSLISAAAGSVGLQALPVGNPSDPSLLIDGTIWECA  
AGDPCDAWCDALISLRAFGYGVDFRLKVDAPKTSVMGAKPTGSAANVTAVD  
RNPAYNKLHDAEFTNAGFLINLWDREFTLGLASNGYLRGNSATNLVGLFV  
KGTYNANELPNVSLNGVVELYTDISFNSVGARGALWEGCATLGAFOYAOQSPK  
VEELNVICNVQSFVNAPKPGYGVAFPLPDAGVATATGKATINHEWQVGLSY  
RLNSLVPYIGVQMSRPTFDADNIRIAQPLPTAVLNLTAWNPSSLGNATLSTDSFS  
DFQIVSCQINKFKSRKACGVTVGATLVDADKSLTAEARLINERAHVSGQPRF"  
1. .1170  
/gene="MOMP"  
70. .1167  
/gene="MOMP"  
/product="major outer membrane protein"  
BASE COUNT 319 a 255 c 247 g 349 t  
ORIGIN

Query Match 99.6%; Score 1096.4; DB 1; Length 1170;  
Best Local Similarity 99.9%; Pred. No. 2.8e-275;  
Matches 1097; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 4 ttgctgtagggaaccccttctgacccaagcttatttaattggttaccatattgggaaggt 63  
Db 70 TTGCTGTAGGGAACCCCTCTGATCCCAAGCTATTAAATGATGTTACATATGGAGGT 129  
Qy 64 gctgcaggagatccttgcgactcttgcctacttgcctgctgagcgtctattagcttaagct 123  
Db 130 GCTGCAGGAGATCCCTTGGCATCTTGGCTACTTGGTGGAGCTATTAGCTTACGTGCT 189  
Qy 124 ggattttacggagactatgttttcgacgtatctttaaagtagatgcacctaataacattt 183  
Db 190 GGATTTTACGGAGACTATGTTTTCGACCGTATCTTAAAGTAGATGCGACCTTAAACATT 249  
Qy 184 tctatggagccaaagcctactggtatccgctgctgcaaaactactactgcogtagataga 243  
Db 250 TCTATGGAGCCAGCCCTACTGATCCGCTGCTGCAAACTACTACTGCGCTAGATAGA 309  
Qy 244 cctaaccggcctacaataaagcattacacgatgcagagtggttgcactaaatgcaggcttc 303  
Db 310 CCTAACCCGGCTACAAATAGCAATTTACAGATGCAGAGTGGTTCACCTAATGAGCTTC 369  
Qy 304 attgcttaaacatttggagcgtcttgcgttctgtacttgcgttgcgttgcgttgcgttgc 363







```

|||||
Db 739 TCTGTAGCTCAAACTACCAAAAGTCTTTTAAAGTAAAGTGTGTAAGTCTAC 798
Qy 484 acagacaccttttcttcttgagcgttagcgtcgtgagcgttattggaatgcggtgt 543
Db 799 ACAGACACCTCTTCTCTGGAGCGTAGGCGCTGCTGGAGCTTTATGGAAATGCGGTGT 858
Qy 544 gcaactttggagcgtgaattccaatatgcacagtcaccaacacaaagttgaagaactta 603
Db 859 GCAACTTTGGAGCGTGAATCCAATATGCACAGTCCAAACCTAAAGTTGAAGAAGTAA 918
Qy 604 gtgacttgaacgtatcgcaattctctgtaaacacaaacccagggcctaataaggcgttgc 663
Db 919 GTGATCTGTACGTAGTCAATCTCTCTAACAACACCAAGGCGTATAAGGCGTGTCT 978
Qy 664 ttcccttggccaacagcgtggcgttagcagcagcagcagcagcagcagcagcagcagc 723
Db 979 TTTCCCTTTGCCAACAGCGTGGCGTAGCAGCAGTACTGGAACAAAGTCTGCGACCATC 1038
Qy 724 aattatcatgaatgcaagtaggagcgtctctctctacagactaaactcttttagtgcca 783
Db 1039 AATTATCATGAATGCAAGTAGGAGCCCTCTAICTTATAGACTAAACTCTTTAGTGCCA 1098
Qy 784 tacattggagtacaaatgtctcgagcaacttttggatgctgataacatccgcattgctcag 843
Db 1099 TACATTGGAGTACAATGTCTCGAGCAACTTTTGTATGCTGATAACATCCGCAATTGCTCAG 1158
Qy 844 ccaaaactactacagcgtgttttaaaactaaactgaatgaagcagcagcagcagcagcagc 903
Db 1159 CCAAACTACCTACAGCTGTTTAAACTTAACTGCATGCAACCCCTCTTTACTAGGAAAT 1218
Qy 904 gccacagcattgtctactactgattcgtctcagacttcagcaaatgtttcctgtcag 963
Db 1219 ACACAACATGGCTACTAGTATGCTGCTCAGACTTCATGCAAAATGTTTCTGTCAG 1278
Qy 964 atcaacaagttaactctagaaaagcgtgtgagttactgttaggagcagcagcagcagcagc 1023
Db 1279 ATCAACAAGTTTAAATCTAGAAAAGCTGTGGAGTACTGTAGGAGCTACTTTAGTTGAT 1338
Qy 1024 gctgataaagtctactactcagaaagcgtgtttaaataacagagagcagcagcagcagcagc 1083
Db 1339 GCTGATAAATGTCTACTTACTCAGAAAGCTGCTTTTAAATACGAGAGAGCTGCTCACGTA 1398
Qy 1084 tctggtcagttcagattc 1101
Db 1399 TCTGGTCAGTTCAGATT 1416

RESULT 9
CHTOMPEQ CHTOMPEQ 1170 bp DNA BCT 04-APR-1994
LOCUS Chlamydia pneumoniae major outer membrane protein gene, complete cds.
DEFINITION
ACCESSION L04982
VERSION L04982.1 GI:289840
KEYWORDS MOMP gene; major outer membrane protein.
SOURCE Chlamydia pneumoniae DNA.
ORGANISM Chlamydia pneumoniae
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 1170)
AUTHORS Storey,C., Lusher,M., Yates,P. and Richmond,S.
TITLE Evidence for Chlamydia pneumoniae of non-human origin
JOURNAL J. Gen. Microbiol. 139, 2621-2626 (1993)
MEDLINE 94103736
FEATURES
source
1..1170
/organism="Chlamydia pneumoniae"
/specific_host="Equus caballus"
/db_xref="taxon:83558"
1..1170
/gene="MOMP"
1..69
/sig_peptide

```

```

/note="putative"
1..1170
/gene="MOMP"
/note="precursor; putative"
/codon_start=1
/transl_table=11
/product="major outer membrane protein"
/protein_id="AA17397.1"
/db_xref="GI:289841"
/translation="MKLLKSLLSAFAAGVSGSLQALPVGNPSPSLLDITWEGA
AGPCDPCATWDAISLRAGFYGDYVDFRLKIDAPKTFSGAKFTGSATANTITAVD
RPNPYNKHLIDAEFTNAGFIALNWDREDFVCTLGASNGYVKGNSAFLNVLFGV
KGTSVANELPNVSLNGVIELYDTTTFAMSVGARGALWECGATGAEQYQAQSKPK
VELNLVICNVOSLKNPKYKGVAFLPTDAGVTAAGTKSATINHEWVGASLSY
RLNSLPYIGVOMSRATFDADNIRIAPKLPTAILNLTAWNPSLIGSATAVSSDDQFS
DFMQIVSCQINKFKSKACGVTGATLVADKWSLAEARLINEARAHISQGRF"
70..1167
/gene="MOMP"
/note="putative"
/product="major outer membrane protein"
312 a 246 c 251 g 361 t
BASE COUNT
ORIGIN
Query Match 90.6%; Score 997.2; DB 1; Length 1170;
Best Local Similarity 94.3%; Pred. No. 2e-249;
Matches 1035; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
Qy 4 ttgctgtagggaacccttctgacccaagcttatttaattgattggtacaaatggaaagt 63
Db 70 TTGCTGTAGGGAACCCCTTCTGATCCAAAGCTTATTAATGTATGTTGATGGAAGGT 129
Qy 64 gctcagagagatccttgcgactccttgcctacttggcgcgacttattagcttgcgtct 123
Db 130 GCTCGAGAGATCCTTGGGATCCTTGCCTACTTGGTCGAGCGCTATTAGCTTACGTCT 189
Qy 124 ggaatttcagagactatgttttcgacgctatcttaaaagttagatgcacctaataatt 183
Db 190 GGATTTTACGGAGACTATGTTTTCGACCGTATCTTAAAGATAGATGCGCCTAAACATTT 249
Qy 184 tctatggagagcaagcctactggatccgctcgtcgaactactactacgctagataga 243
Db 250 TCTATGGAGGCAACACTACAGGATCCGCTACTTGGCAACTACTACCGCGGTAGTAGA 309
Qy 244 cctaacccgctcacaataaagcatttacacagatgcagagtgttgcactaaatgcaggttc 303
Db 310 CCTATCTGCTGCTACAAATGAAGCAITATACGATGCAGATGGTTTACATTAATGCAGGCTC 369
Qy 304 attgcttaaacatttgggagcgttcttgcgttcttctgtacttttaggagcgtcttaattgt 363
Db 370 ATTGCTTTAAACATTTGGGATCGCTTTGATGTTTCTGTACCTTGGAGCTTCTAATGCT 429
Qy 364 tacattagagaaactctacagcgttcaatcctcgttgcgtttatttcgaggttaaaagtact 423
Db 430 TACGTTAAAGGAAACTCTGCGCGCTCAATCTCGTTGGTTTATTTCGGAGTTTAAAGGTACT 489
Qy 424 actgtaaatgcaaatgaactacaaacgctttctttaaagtaacgaggtgttgaactttac 483
Db 490 TCTGTAATGCAATGAAGTCACTACCAACCGTTCTTTAAGTAATGGAGTATTCGAGCTCTAT 549
Qy 484 acagacaccttttcttcttgagcgttagcgtcgtgagcgttattggaatgcggttgc 543
Db 550 ACAGATACTACTCTTTGCTTTGGAGCGTAGTGTCTGCTGGAGCTTTTATGGAAATGCGGTGT 609
Qy 544 gcaactttgggagcgtgaattccaatatgcacagtcaccaacacaaagttgaagaactta 603
Db 610 GCCACTTTGGGAGCTGAATTCGAATATGCACAGTCTAAACCTTAAAGTTGAAGAAGTAA 669
Qy 604 gtgacttgaacgtatcgcaattctctgtaaacacaaacccagggcctaataaggcgttgc 663
Db 670 GTGATCTGTAGTATCGCAATCTCTTAAATAAACCAAGGCGTATAAAGCGGTGCT 729
Qy 664 ttcccttggccaacagcgtggtcgttagcagcagcagcagcagcagcagcagcagcagc 723

```







Qy	475	gaacttacacagacacctcttctcttgagctgagcgctcgtgagccttatgggaa	534
Db	241	GAACCTTTACAGACACACCTCTTTCTCTGGAGCGTAGGCGCTCGTGAGCCTTATGGAA	300
Qy	535	tgcggttgcaacttgggagctgaattccaatatgcacagtcctcaaaccttaaaagttaa	594
Db	301	TGCGGTTGTGCAACTTTGGGAGCTGAATTCGAATATGCACAGTCCAAACCTAAAGTTGAA	360
Qy	595	gaacttaagtgtatctgttaacgtatcgcaattctctgttaaaacccaaagggtataaa	654
Db	361	GAACCTTAATGTGATCTGTGAAGCTATCGCAATCTCTGTAAACAAACCCCAAGGGGTATAA	420
Qy	655	ggcgttgcttccctctgcacaacagcctggctagcaacagctacttggaacaaaagtct	714
Db	421	GGCGTGTGCTTTCCCTCTGCCAACAGACGCTGGCGTAGCAACAGCTACTTGGAAACAAAGTCT	480
Qy	715	gcgaccatcaattatcatgaatggcgaagttagagcctctctatcttacagactaaactct	774
Db	481	CGGACCATCAATTAATCATGAATGGCAAGTAGGAGCGCTCTCTATCTTACAGACTAAACTCT	540
Qy	775	ttagtgcatacattggagtcacaaatggtctcgagcaacttttgcgtgcgtataaacatcgc	834
Db	541	TTAGTGCCATACATTTGGAGTACATGGTCTCGAGCAACTTTTGATGCTGATAACATCGCG	600
Qy	835	attgctcagccaaactaccctcacagctgttttaaaacttaactgcattggaacccctcttta	894
Db	601	ATTGCTCAGCCAAACTACTTACAGCTGTTTAAACTTTAATGCATGCAACCCCTCTTTA	660
Qy	895	ctaggaatg 904	
Db	661	CTAGGAATG 670	
RESULT	12		
AF131229			
LOCUS	670 bp	DNA	BCT
DEFINITION	Chlamydomophila pneumoniae mutant major outer membrane protein (MOMP)		01-JUN-1999
ACCESSION	AF131229	gene, partial cds.	
VERSION	AF131229		
KEYWORDS	AF131229.1	GI:4928267	
SOURCE	Chlamydomophila pneumoniae.		
ORGANISM	Chlamydomophila pneumoniae		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.		
AUTHORS	1 (bases 1 to 670)		
TITLE	Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Bing,L.-M.		
JOURNAL	Presence of viable Chlamydia pneumoniae in fetal calf serum and in epithelial-derived cell lines		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 670)		
TITLE	Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Bing,L.-M.		
JOURNAL	Direct Submission		
FEATURES	Submitted (25-FEB-1999) Pathology, Vanderbilt University, C-3321 Medical Center North, Nashville, TN 37232, USA		
source	Location/Qualifiers		
	1..670		
	/organism="Chlamydomophila pneumoniae"		
	/db_xref="taxon:83558"		
	/note="synonym: Chlamydia pneumoniae"		
	isolated from cryptically-infected H292 cells"		
gene	<1..>670		
	/gene="MOMP"		
CDS	<1..>670		
	/gene="MOMP"		
	/note="similar to MOMP gene sequences of Chlamydia pneumoniae presented in GenBank Accession Numbers M64064, M34942, and M64063"		
	/codon_start=1		
	/transl_table=1		
	/product="mutant major outer membrane protein"		
	/protein_id="AAD33510.1"		
	/db_xref="GI:4928268"		
	/translation="VDRPNAAYNMHLDAEWFNTAGFALNIWDRFVCTLGASNG"		







Db	58	TTGCCGTGTAGGGACCCAGCTGAACCAAGTTTATTAATCGATGGCACTATTGTGGGAAGGT	117
Qy	64	gctgcaggagatccttgatcgatcctgcgactctggtgcagcgtattagcttaccgtct	123
Db	118	GCATTAGGAGATCCTTGGCATCCTTTGCCTACTTTGGTGTGACGCCATTAGCATCCGCGCA	177
Qy	124	ggattttacggagactatgttttcagccgtatctttaaagttagatgcacotcaaaacattt-	182
Db	178	GGATACTACGGAGATTATGTTTTCGATCGCTGTATTAAGAAGTTGATGTAATAAAACTATC	237
Qy	183	--ttctatggagccaagcctactcgtatccgctcgtcgaactatactactcgcgtagat	240
Db	238	AGCGGTATGGGTGCAGCTCCTTACAGAAAGCCGACGAGCCGGATTACAAAACTCTCAGAT	297
Qy	241	agacctaaaccggcctacaataaagcatttacacgatcagagtggttcacataatcgaagc	300
Db	298	AGACCAACATCGCTTATGGCAACACATTGCAAGACGCTGAGTGGTTTACGAATGCGAGCT	357
Qy	301	ttcattgccttaaaacatttggtgatcgctttagtgttttctgtactcttaggagcttcta	360
Db	358	TTCTCGCATTAATATCTGGGATCGCTTGATATTTCTGCACATTAGGTGCTTCCAAT	417
Qy	361	ggttacattagagaacctctacagcgtttcaatcgttggtttattcggagttaaaggt	420
Db	418	GGGTACTTCAAGCTAGTTCTGCTCAATCAACCTCGTTGCTTTGATTGGTGTAAAGGA	477
Qy	421	actactgtaaatgcaaatgaactacaaacagtttcttttaagtaacgagttgttgaaact	480
Db	478	ACCTCGTAGACGCTGATCAACTTCCAAGCTAGGCACTCACTCAAGGTATTTGTAGTTT	537
Qy	481	tacacagacacctctctctggagcgttagcgcctcgtggagccttatgggaatcggt	540
Db	538	TACACAGATACACATCTCTTGGAGCTAGGTGCACGTGGTCTTTATGGGAATGNGGT	597
Qy	541	tgtgcaactttgggagctgaattccaatatgcacagtcaccaacccctaaagttgagaact	600
Db	598	TGTGCAACTTTAGGAGCTGAATTCAGTATGCTCAATTAATCTCAAAATTTGAATGCTG	657
Qy	601	aattgactctgaacgtatcgaacttctctgaacaaacccagggctataa-----a	654
Db	658	ANTGTAATCTCCAGCCCCACANNTTGTAGTTTCAAGCCTTAGAGATACAGGGGAACA	717
Qy	655	ggcgtgtcttcccttgcacacagcgtcgtcgtagcaacagctactggaaacaaagct	714
Db	718	GGATCGAACTTTCCTTACTCTTAACAGCTGTGCAGATGGTCTACAGATACTAAATCT	777
Qy	715	gcgacataaattatcatgaatggaagtagagcctctctatttcacagactaaactct	774
Db	778	GCAACATCAAAATATCATGAATGGCAAGTTGGTTTAGCGCTCTCTTACAGATTGAACATG	837
Qy	775	ttagtgcatcacattggagtacaatggtctcgcagcaacttttgatcgtgataacatcgc	834
Db	838	CTTGTCTTACATTGGCGTAAACTGGTCAAGACAACTTTTGTATGCTGACTCTATCCG	897
Qy	835	attgctcagcacaatactacagctgttttaacttaactgaactgcatggaaacctcttta	894
Db	898	ATCCTCAACCTTAAATTAGCCGCTGCTGTTTGAACCTTGACCATGGAACCCAACTCT	957
Qy	895	ctaggaaatgccacagcattgtctactactgattcgttctcagactctcatgaattggt	954
Db	958	TTAGGGGAAGCTACAGCTTTAGATGCTACCAACAAATTCGCGACTTCTTACAAATCGCT	1017
Qy	955	tctcgtcagatcaacaagtttaacttagaagaagcttggtgagttactcttaggagctact	1014
Db	1018	TCGATTTCAGATCAACAAATGAATCTAGAAAAGCTTGTGTTGACGTGTTGGTGCAACG	1077
Qy	1015	ttagttgatcgtataaattggtcaactctcagagaagctcgttttaattcaacgcagagact	1074
Db	1078	TTAATCAGCGCTGACAAATGGTCAATCACTGGTGAAGCAGCGCTTAATCAATCAAGAGCT	1137
Qy	1075	gctcacgtatctgggttcagttcagattc 1101	
Db	1138	GCTCATGATGATGCTCAATTCAGATTC 1164	

RESULT	14
AF272945	1170 bp DNA BCT 25-JAN-2001
LOCUS	Chlamydomophila abortus major outer membrane protein precursor (Ompl)
DEFINITION	gene, complete cds.
ACCESSION	AF272945
VERSION	AF272945.1 GI:12483891
KEYWORDS	Chlamydomophila abortus.
SOURCE	Chlamydomophila abortus
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE	1 (bases 1 to 1170)
AUTHORS	Vretou,E., Psarrou,E., Kaisar,M., Vlisidou,I., Salti-Montesanto,V. and Longbottom,D.
TITLE	Identification of protective epitopes by sequencing of the major outer membrane protein gene of a variant strain of Chlamydia psittaci serotype 1
JOURNAL	Infect. Immun. 69 (1), 607-612 (2001)
MEDLINE	20569239
PUBMED	11119563
REFERENCE	2 (bases 1 to 1170)
AUTHORS	Longbottom,D. and Vretou,E.
TITLE	Direct Submission
JOURNAL	Submitted (29-MAY-2000) Moredun Research Institute, Bush Loan, Penicuik EH26 0PZ, UK
FEATURES	Location/Qualifiers
source	1..1170
	/organism="Chlamydomophila abortus"
	/isolate="LLG"
	/specific_host="Capra aegagrus hircus"
	/db_xref="taxon:83555"
	/country="Greece"
sig_peptide	1..66
	/gene="Ompl"
CDS	1..1170
	/gene="Ompl"
	/note="MOMP; porin"
	/codon_start=1
	/transl_table=11
	/product="major outer membrane protein precursor"
	/protein_id="AAC53881.1"
	/db_xref="GI:12483892"
	/translation="MKKLLKSALLFAATGSALSLOALPVGNPAEPSSLIDGTWEGAS GDPCDPCSTWCDAISIRAGYGYGVDFDRVLKVDNKTITGMGAVPVTGTAADAKYRTDT RPNLAYGHKLQDAEWETNAAFALNLTWRDFDICTLGASNGYFKASSAANFLVGLIGV KGSVAADOLPNVGJTGQIVVEFTYDTTFSWSVGARGALWECGCATLGAEFOYAQSNNPK IEMLVNSPPAQSVVHKPRGYKKTFPLPLTAGCTDOATDTKSATIKYHEWQVGLALSY KLMNLVPYICNVNSRAFTRDADAIRIAQPKLAAAVLNLTWNPTLLGEATTLDTNNKFA DFQIQAISIQINKMRKACGVAVGATLIDADKSWITGEARLINERAAHMAQAQFRF"
gene	1..1170
	/gene="Ompl"
mat_peptide	67..1167
	/gene="Ompl"
	/product="major outer membrane protein"
BASE COUNT	322 a 263 c 255 g 330 t
ORIGIN	
Query Match	54.5%; Score 600.2; DB 1; Length 1170;
Best Local Similarity	72.2%; Pred. No. 6.6e-146;
Matches	795; Conservative 0; Mismatches 303; Indels 3; Gaps 1;
Oy	4 ttgccttagggaaaccccttcgatccaagccttattaatgcatgtgtacaatatgggaaggt 63
Dd	
	67 TTGCTGTAGGGAAGCCCAAGCTGAACCAAGTTTATTAAATCGATGGCACCATTGTGGGAAGT 126
Oy	64 gctcacagagactcttgcgactcttgccacttgcttggtgcagcctattaggcttacgtgt 123
Dd	
	127 GCTTCAGGTGATCCTTGTGCATCTTGTCTACTTGGTGTGTGCTATCAGCATCGCGCA 186
Oy	124 ggattttacggagactatgttttcgaccgactctcttaaaagttagatgcgccataaacattt 183



Db 187 GGATACACGAGATATGTTTCGATCGTGTATTAAAGTTGATGAATAAACTATC 246  
 Qy 184 tc---tatggagcaagcctactgagtcgctgctgcaaaactatactactgagcgtatag 240  
 Db 247 ACCGGCATGGGTGCGATGTTCTTACAGGAACCGGAGCAGCTGATATACAAACTCCTACGGAT 306  
 Qy 241 agacctaaacccgctacataaagcatttacacgctgagagtggttcaactaatcagc 300  
 Db 307 AGACCCAAACATCGCTTACGGCAACACTTACAGAGCGCGAATGTTTACCAGTACGCT 366  
 Qy 301 ttaatgcttaaaacatttgggagcgtttgatgttttctgtacttttggagccttctaat 360  
 Db 367 TTCTCGCATTAATATCTGGATCGCTTGTATATTTCTGCACATTAGCGCTTCTTAAT 426  
 Qy 361 gttacattaggaagaaactacagcgttcaatctcgttggttttattcggagttaaagt 420  
 Db 427 GGGTACTTCAAGCTAGTGTCTCGGCATTTCAACCTCGTGGTGTGATGGTGTAAAGGA 486  
 Qy 421 actactgtaaatgcaaatgaaactaccaaagcgtttctttaagtaacgaggtttgtaactt 480  
 Db 487 TCCTCGTAGCAGCTGATCAGCTTCCCAATGTAGGCATCACCTCAAGGAATCGTTGAATTT 546  
 Qy 481 tacacagacacctcttctcttgaggcgttaggcgtctgtagccttattgggaatcggt 540  
 Db 547 TATACAGATACAACTATCTTGGAGTGTAGGTGCACGCGAGCTTTATGGAGTGTGT 606  
 Qy 541 tttgcaactttggagctgaattccaatatgcacagtcacacgttcaaaccttaagttgaagaactt 600  
 Db 607 TGTGCAACTTAAAGAGCAGAGTCCAACTAGCTCAGTCTAATCTTAAATTAATTAATGTTG 666  
 Qy 601 aatgtgactgttaacgtatcgcaattctctgttaaacacacccaaagggctataaagcggt 660  
 Db 667 AATGTAGTCTCCAGCCAGCAGCAATTTGTGGTTTCAAGCCCTAGAGGATACAAAGGAACA 726  
 Qy 661 gtttccctctgcaacagacgtcgttagcagcaacagctactgtggaacaaagctcgcgacc 720  
 Db 727 GCAATTCCTTACCTCTACAGCTGGTACTGATGACGGAACCTGACACTAAGTTCGGCTACA 786  
 Qy 721 atcaattatcatgaatggcgaagtaggagcctctctctatcttctacagactaaactcttagtg 780  
 Db 787 ATTAATAACCAAGTGAATGCAAGTGTGTTAGCGCTCTCTTATCGATTAAGATGCTTGT 846  
 Qy 781 ccatacattggagtgcaatggtctgagcaacttttgatgctgataacatccgcattgct 840  
 Db 847 CTTTACATTTGGGCTAACTGTTGTCAGCAGCACTTTTGTGCTGTCGCTATCCGCTCGCT 906  
 Qy 841 cagccaaactacacagctgttttaaaacttaactgaatgcaaccccttcttactagga 900  
 Db 907 CAACCTAATTAAGTGTCTGTGTTAACTTTGACACCATGGAACCCCAACCCCTTTTAGGG 966  
 Qy 901 aatgccaacagcattgtctactactgattcgtttctcagacttcattgcaaatgtttcctgt 960  
 Db 967 GAAGCTACAACTTATAGATTAACCAACAAATTCGCTGACTTCTGCAATTCGCTCGATT 1026  
 Qy 961 cagatcaacaagtttaactatagaaaagctgtgaggtactgttaggagcactttagtt 1020  
 Db 1027 CAGATCAACAAATGAAGTCTAGAAAAGCTTGTGGTGTAGCTGTTGGTGAACGTTAATC 1086  
 Qy 1021 gatcgtgataaagctactacagagctcgttttaataataacagagagagcgtctcac 1080  
 Db 1087 GAGCTGACAAATGCTAATCTACTGTAAGCAGCCTTAATCAATGAAGAGCGCCTCAC 1146  
 Qy 1081 gtatcgtggtcaggttcagattc 1101  
 Db 1147 ATGAATGCTCAATTCAGATTC 1167

RESULT 15  
 CHTOMPAAD 1261 bp DNA BCT 07-MAY-1999  
 LOCUS Chlamydomophila abortus strain B577 major outer membrane protein  
 DEFINITION precursor (ompA) gene, complete cds.

ACCESSION M73036  
 VERSION M73036.1 GI:144564  
 KEYWORDS Chlamydomophila abortus.  
 SOURCE Chlamydomophila abortus  
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia  
 REFERENCE 1 (bases 1 to 1261)  
 AUTHORS Kaitenboeck, B., Kousoulas, K.G. and Storz, J.  
 TITLE Structures of and allelic diversity and relationships among the major outer membrane protein (ompA) genes of the four chlamydial species  
 JOURNAL J. Bacteriol. 175, 487-502 (1993)  
 MEDLINE 93123168  
 FEATURES Location/Qualifiers  
 source 1..1261  
 /organism="Chlamydomophila abortus"  
 /strain="B577"  
 /specific\_host="Ovis aries"  
 /db\_xref="taxon:8355"  
 /note="submitted as Chlamydia psittaci"  
 80..1246  
 /gene="ompA"  
 80..1249  
 /codon\_start=1  
 /transl\_table=11  
 /product="major outer membrane protein precursor"  
 /protein\_id="AAD29102.1"  
 /db\_xref="GI:4757612"  
 /translation="MKLLKSLALLFAATGTSALSLSLQALPVGNAPSPSLIDGTMMEGAS  
 GDCPCSTWCDAISIRAGYGDYDFVRVLRKVDNKTITGMGAVTGPAAAYKTPTD  
 RPNYAGKHLQDAEMFTNAFALININDRFIDFCTLGASNGYFKASAFNLVGLIGV  
 KGSSIAADQLPNVGITGIVFYDTTTPSVSGARGALWECGCATLGAFFYAOASNPK  
 IEMLVYSSPAOFVVKPRGKGTAFPLPTAGTDOATDTSATIKYHEMOVGLALSY  
 RLNLVPYISVNSRATFDADAIPIAOPKLAANLNTWNPFTLLGEGATLDTSNKPA  
 DFLOIASIQINKMSRKACGVAVGATLIDADKNSITGEARLINERAHMAHQFRP"  
 BASE COUNT 359 a 275 c 277 g 350 t  
 ORIGIN  
 Query Match 54.5%; Score 600.2; DB 1; Length 1261;  
 Best Local Similarity 72.2%; Pred No. 6.e-146;  
 Matches 795; Conservative 0; Mismatches 303; Indels 3; Gaps 1;  
 Qy 4 ttgctgttagggaaccccttctgataccagcttataattgattggtacaaatgggaagt 63  
 Db 146 TTGCTGTAGGGAACCCAGCTGAACCAAGTTTATTATCGATGGCAGCTATGTGGGAAGT 205  
 Qy 64 gctgcagagatccttgcgactcttgcctactgtgtgcgcgacttagcttagctgct 123  
 Db 206 GCTTCAGGTGATCTCTGGGATCTTGTCTACTTGTGTGTATCAGCATCGCGCA 265  
 Qy 124 ggattttacgagactatgttttcgacctatcttaaaagttagtgacacctaataattt 183  
 Db 266 GGATTAAGGAGATATGTTTTCGATCGTGTATTAAGATTTGATGAATAAACTATC 325  
 Qy 184 tc---tatggagcaagcctactgagtcgctgctgcaaaactatactactcgcgtatag 240  
 Db 326 ACCGCGATGGGTGCGATTCCTTACAGGAACCGCAGCAGCTAATTACAAACTCCTACGAT 385  
 Qy 241 agacctaacccgctacataaagcatttacacgctgagagtggttcaactaatcagc 300  
 Db 386 AGACCCAAACATCGCTTACGGCAACACTTACAGAGCGCGAATGTTTACCAGTACGCT 445  
 Qy 301 ttaatgcttaaaacatttgggagcgtttgatgttttctgtacttttggagccttctaat 360  
 Db 446 TTCTCGCATTAATATCTGGATCGCTTGTATATTTCTGCACATTAGCGCTTCTTAAT 505  
 Qy 361 gttacattaggaagaaactacagcgttcaatctcgttggttttattcggagttaaagt 420  
 Db 506 GGGTACTTCAAGCTAGTGTCTCGGCATTAACCTCGTGGTGTGATGGTGTAAAGGA 565  
 Qy 421 actactgtaaatgcaaatgaaactaccaaagcgtttctttaagtaacgaggtttgtaactt 480







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 20:17:12 ; Search time 685.35 Seconds  
(without alignments)  
1377.275 Million cell updates/sec

Title: US-09-391-606-14

Perfect score: 1101

Sequence: 1 atgttcctgtagggaacc.....tatctgtcagttcagattc 1101

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

- 1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	600.2	54.5	1261	20 AAX25047	Chlamydia psittaci
2	526.4	47.8	1660	20 AAX25048	Chlamydia psittaci
3	523.2	47.5	1209	20 AAX25046	Chlamydia psittaci
4	514.6	46.7	1578	21 AAX64764	C. pneumoniae sero
5	514.6	46.7	1578	22 AAX56267	Chlamydia trachoma
6	479.4	43.5	3133	19 AAX62447	Chlamydia trachoma
7	479.4	43.5	3133	21 AAX292753	DNA encoding Chlam
8	477.8	43.4	3133	19 AAX40646	Chlamydia trachoma
9	474.6	43.1	3133	7 AAN60007	Sequence encoding
10	397.2	36.1	1038602	20 AAX201425	Complete genome se
11	352.2	32.0	744	20 AAX25045	Chlamydia psittaci

12	328.8	29.9	1047	21 AAX08124	C. trachomatis MOM
13	321.8	29.2	726	20 AAX25044	Chlamydia psittaci
14	287.4	26.1	1362	21 AAX08122	C. trachomatis MOM
15	287.4	26.1	1371	21 AAX08120	C. trachomatis MOM
16	287.4	26.1	1542	21 AAX08123	C. trachomatis MOM
17	254	23.1	1452	21 AAX08121	C. trachomatis MOM
18	205.2	18.6	831	21 AAX08125	C. trachomatis MOM
19	140	12.7	720	7 AAN60006	Sequence encoding
20	140	12.7	720	19 AAX62446	Vector lambda gtl1
21	140	12.7	720	19 AAX40645	Chlamydia trachoma
22	140	12.7	720	21 AAX292752	DNA encoding Chlam
23	87.6	8.0	936	22 AAF58252	Oligonucleotide D1
24	87.6	8.0	936	22 AAF58254	Oligonucleotide D1
25	87.6	8.0	936	22 AAF58257	Oligonucleotide D1
26	87.6	8.0	936	22 AAF58259	Oligonucleotide D2
27	87.6	8.0	936	22 AAF58262	Oligonucleotide D2
28	87.6	8.0	936	22 AAF58255	Oligonucleotide D1
29	87.2	7.9	936	22 AAF58252	Oligonucleotide D1
30	87.2	7.9	936	22 AAF58254	Oligonucleotide D1
31	87.2	7.9	936	22 AAF58257	Oligonucleotide D1
32	87.2	7.9	936	22 AAF58259	Oligonucleotide D2
33	87.2	7.9	936	22 AAF58262	Oligonucleotide D2
34	87.2	7.9	938	22 AAF58255	Oligonucleotide D1
35	75	6.8	259	15 AAX54801	Chlamydia OMP frag
36	71.6	6.5	259	15 AAX54800	Chlamydia OMP frag
37	54.6	5.0	245	15 AAX54798	Chlamydia OMP frag
38	54.6	5.0	246	15 AAX54799	Chlamydia OMP frag
39	47	4.3	50	20 AAX99685	Chlamydia OMP ge
40	47	4.3	50	20 AAX99686	Chlamydia OMP ge
41	45.4	4.1	50	20 AAX99688	Chlamydia OMP ge
42	43	3.9	75	19 AAX56996	C. trachomatis maj
43	43	3.9	75	20 AAX61262	Major outer membra
44	43	3.9	75	21 AAX08129	MOMP containing fu
45	41.2	3.7	80	14 AAX43617	Chlamydia trachoma

## ALIGNMENTS

RESULT 1  
AAX25047  
ID AAX25047 standard; DNA; 1261 BP.  
XX  
AC AAX25047;  
XX  
DT 05-JUL-1999 (first entry)  
XX  
DE Chlamydia psittaci major outer membrane protein DNA.  
XX  
KW Major outer membrane protein; MOMP; psittacosis; infection;  
KW vaccine; genetic immunisation; ss.  
XX  
OS Chlamydia psittaci.  
XX  
FH Key Location/Qualifiers  
CDS 80..1249  
FT /\*tag= a  
XX  
PN W09910005-A1.  
XX  
PD 04-MAR-1999.  
XX  
PF 28-AUG-1998; 98WO-US17943.  
XX  
PR 28-AUG-1997; 97US-0057147.  
XX  
PA (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.  
XX  
PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
XX  
DR WPI: 1999-254214/21.  
XX  
P-PSDB: AAX98188.



PT A new vaccine for Chlamydia psittaci infections  
 XX Disclosure; Page 57-60; 72pp; English.

XX This DNA sequence codes for the major outer membrane protein (MOMP, see AAW98187) of Chlamydia psittaci strain B577. A claimed MOMP polypeptide (see AAW98184) comprises regions VD3 and VD4 of MOMP, i.e. it lacks regions VD1 and VD2. A claimed vaccine composition includes the MOMP polypeptide, optionally fused to a maltose binding protein. Also claimed are an isolated nucleic acid encoding the polypeptide, a vector, and a method of preventing C. psittaci infection by administering the vaccine containing the MOMP polypeptide. Vectors encoding MOMP polypeptides lacking regions VD1 and VD2 are useful for genetic vaccination. The vaccines are used to prevent C. psittaci infection, especially in birds.

XX Sequence 1261 BP; 359 A; 275 C; 277 G; 350 T; 0 other;

Query Match 54.5%; Score 600.2; DB 20; Length 1261;  
 Best Local Similarity 72.2%; Pred. No. 5.1e-173;  
 Matches 795; Conservative 0; Mismatches 303; Indels 3; Gaps 1;  
 QY 4 ttgctgtagggaaccccttctgatcccaagcttatttaattggtgtacaaatgggaaggt 63  
 Db 146 ttgctgtagggaacccagcgaacaaagtttatttaattggtgtacaaatgggaaggt 205  
 QY 64 gctgcagagagactctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 123  
 Db 206 gctgcagagagactctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 265  
 QY 124 ggaatttccagagactgctgctgctgctgctgctgctgctgctgctgctgctgctgct 183  
 Db 266 ggaatttccagagactgctgctgctgctgctgctgctgctgctgctgctgctgctgct 325  
 QY 184 tc---tatggggcccaagcctactgctgctgctgctgctgctgctgctgctgctgctgct 240  
 Db 326 accgcatgggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 385  
 QY 241 agactaaccggcctacaaataagacatttacagatgcagagtggttccactaactcagcgc 300  
 Db 386 agaccacaacatgcttaccgcaacacacttacagagcgcgaatggttccactaactcagcgc 445  
 QY 301 ttcatggcttaaacatttggtgctgctgctgctgctgctgctgctgctgctgctgctgct 360  
 Db 446 ttctcgcattgaatctggtgctgctgctgctgctgctgctgctgctgctgctgctgct 505  
 QY 361 ggttacattaggaagaaactctacagcttcaatctctgctgctgctgctgctgctgctgct 420  
 Db 506 ggggtactcaagactagttctcgccattcaacctcgttgggttattggtggttaaaagga 565  
 QY 421 actactgtaaatgcaaatgaactacaaacgtttctttaaagtaacggagttgttgaactt 480  
 Db 566 tctccatagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 625  
 QY 481 tacacagacaccttctcttctgagcgtgagcgtgctgctgctgctgctgctgctgctgct 540  
 Db 626 tatacagatacaacattctctgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 685  
 QY 541 tgtgcaacttggagctgaattcccaatatgcacagctccaaaccttaaaagttggaagactt 600  
 Db 686 tgtgcaacttggagcagagttcccaatatgcacagctccaaaccttaaaagttggaagactt 745  
 QY 601 aatgtgactgttaacgtatcgcaattctctgttaacaaaccccaagggctataaagcgtt 660  
 Db 746 aatgtgactgtccagccagcaaatgtgttccaaagccttagaggtacaaaggggaaca 805  
 QY 661 gctttcccttgcacacagcgtgctgagcaacagctactgagcaacaaagcttcgcgacc 720  
 Db 806 gcaattcccttacccttaacagctggtactgacgagcaactgacactaagtcggtacac 865  
 QY 721 atcaattatcatgaatggcaagtaggagcctctctcttcttcttcttcttcttcttcttct 780

Db 866 attaaataccacgaatggcaagttgttttagcgtctcttcttcttcttcttcttcttctt 925  
 QY 781 ccatacattgagtagacaatgggtctgcagaaactttttagctgatacaacatccgattgct 840  
 Db 926 ccttaccattagcgttaactggttcacagcgaactttttagctgatacaacatccgattgct 985  
 QY 841 cagcacaactacactacagctgttttaaaacttaactgcagtggaaccccttcttactagga 900  
 Db 986 caactaaattagctgctgctgtgttaaaactgaccacatggaaccccttcttactagga 1045  
 QY 901 aatgccacagactgttctactactgctgttcttcttcttcttcttcttcttcttcttctt 960  
 Db 1046 gaagctacagcttttagatactacgacaatactgctgacttcttgcgaattgcttcgatt 1105  
 QY 961 cagatcaacaagtttaatacttagaaaaagcttggagttactgtactgagagctactttgatt 1020  
 Db 1106 cagatcaacaatacgaagctcagaagccttgggtgtgctgtgtgtgtgtgtgtgtgtgt 1165  
 QY 1021 gatgctgataatggtcacttactgcagaagcctgtttaaataacagagagctgctcacc 1080  
 Db 1166 gacgtgacaatggttcaatcactggtgagcagcgttaataatgaagagcgcctcac 1225  
 QY 1081 gtatctggtcagttcagattc 1101  
 Db 1226 atgaatgctcaattcagattc 1246

RESULT 2  
 AAX25048  
 ID AAX25048 standard; DNA; 1660 BP.  
 XX  
 AC AAX25048;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE Chlamydia psittaci major outer membrane protein DNA.  
 XX  
 KW Major outer membrane protein; MOMP; psittacosis; infection;  
 KW vaccine; genetic immunisation; ss.  
 XX  
 OS Chlamydia psittaci.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 364..1572  
 FT /\*tag= a  
 XX  
 PN WO9910005-A1.  
 XX  
 PD 04-MAR-1999.  
 XX  
 PF 28-AUG-1998; 98WO-US17943.  
 XX  
 PR 28-AUG-1997; 97US-0057147.  
 XX  
 PA (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.  
 XX  
 PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
 XX  
 DR WPT; 1999-254214/21.  
 DR P-PSDB; AAW98189.  
 XX  
 PT A new vaccine for Chlamydia psittaci infections  
 XX  
 PS Disclosure; Page 62-65; 72pp; English.  
 XX  
 CC This DNA sequence codes for the major outer membrane protein (MOMP, see AAW98187) of Chlamydia psittaci strain 6BC. Claimed MOMP polypeptides (see AAW98183 and AAW98184) comprise regions VD3 and VD4 of an MOMP, i.e. they lack regions VD1 and VD2. Claimed vaccine compositions include such MOMP polypeptides, optionally fused to a maltose binding protein. Also claimed are isolated nucleic acids encoding the polypeptide, a vector, and a method of preventing C.



CC psittaci infection by administering the vaccine containing the  
CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking  
CC regions VD1 and VD2 are useful for genetic vaccination. The  
CC vaccines are used to prevent *C. psittaci* infection, especially in  
CC birds.

Sequence 1660 BP; 492 A; 342 C; 338 G; 488 T; 0 other;

```
Query Match          47.8%; Score 525.4; DB 20; Length 1660;
Best Local Similarity 69.5%; Pred. NO. 2.3e-150;
Matches 792; Conservative 0; Mismatches 306; Indels 42; Gaps 4;
```

Qy 4 ttgccttagggaaaccttctgatcccaaggttattaattgatggtaacaattgggaaggt 63  
|||||  
Dd 430 ttgccttagggaaacccagctgaaccaagttttaatacctatggcaactatgggaaggt 489

**Qy** 64 gctgcaggagatccttgcgataccttcgcactcttggtgcgacgctattagcttacgtgct 123  
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||  
**Db** 490 gcttcaggagatccttgcgataccttcgcactcttggtgcgacccatttagcatccgcqca 549

Oy 124 ggattttacggagactatgttttcgaccgtatctttaaagttagatgcacctaaacattt 183  
||||| ||||| ||||| ||| ||||| ||| ||||| |||  
Db 550 ggaatacaggagattatgttttcgatcggtatttaaaaagtgatgtaataaacattt 609

Qy 184 tctatggagccagcctactggtatccgctgctgcaactatactact----- 231  
Db 610 agcggcatggctgcaactcctacgaggtacaggtacagcgaagtaataactaatcagcca 669

Qy 232 ---gccgtagatgacctaacccggcctacataagaatttcacagatgcagagtgttc 288  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 670 gaagcaaatggcacaccgaacatcgcttacggaaggcatatgcaagatgcagagtgttt 729

**Qy** 289 actaatgagggcttcattgccttaacattgggatcgcttgatgttttctgtacttta 348  
||||| |||| | ||||||||| ||||| ||||| ||||| |||||  
**Db** 730 tcaaatgcagccttccttagccttaaacatttgggatcgcttcgacattttctgcacctta 789

[illegible]

**Qy** 409 ggaggttaaggT-----actactgtaaatagcacaaatgaactaccacaaagtttct 456  
|| || || |  
**Db** 850 ggggtttcagctgcaagtcgaattcttaccgatctccaatgcaacttccctaagtaggc 909  
||||| | ||| |||| | |||||

**Qy**    457   ttaagtaacggagtgtgtgaactttcacacagacaccttcttcttgaggcgtaggcgct   516  
         | | | | | | | | | | | | | | | | | | | | | | | |  
**Db**     910   attaccaagggtgtgtggaatttatatacagacacatcatcttcttgaggcgtaggtgca   969

[illegible]

**Oy** 577 tccaaacctaaagtgaagaactaatgtgatctgtaacgtatcgcaattctcgtaac 630  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
**Dd** 1030 tttaatcctaagattgaaatgctcaacgctcaacttcaggcccagcacaaatttggattcac 1080

Uy 63/ aaacccaaggcgcataaaagcgct-----gcttccctcttgcaaacagacgcctggcgcta 690  
 ||||| | ||||| ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |  
 Db 1090 aaaccaagagcgcataaaaggagctgctcgaaatttctctttaccctataacgcgtcggaaca 114

[illegible]

Db 1210 gcctgtctacagattgaatatgctgttccatatattggcgtaaactgggtcaagagca 1260

1270 acctttgatgctgatactactccgcattgcacacctaataattaaatcgagattctttaac 132

[illegible]

RESULT 3  
AAX25046  
ID AAX25046 standard: DNA: 1209 BP.

AA	
AC	AAX25046;
XX	
DT	05-JUL-1999 (first entry)

DE Chlamydia psittaci major outer membrane protein DNA.  
XX  
KW Major outer membrane protein; MOMP; psittacosis; infection;

XX vaccine, genotype immunization, 55.  
XX  
OS Chlamydia psittaci.  
XX

XX  
XX  
PD  
XX  
04-MAR-1999.

XX  
PR 28-AUG-1997; 97US-0057147.  
XX

XX  
PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
XX

DR P-PSDB; AAW98187.  
XX  
PT A new vaccine for Chlamydia psittaci infections

PS Disclosure; Page 53-55; 72pp; English.  
XX  
CC This DNA sequence codes for the major outer membrane protein (MOMP,

CC isolate (the MOMP gene sequence of this isolate is identical to  
CC that of *C. psittaci* Avian Type C). A claimed MOMP polypeptide (see  
CC AA#98183) comprises regions VD3 and VD4 of MOMP, i.e. it lacks  
CC regions vnt1 and vnt2. A claimed vaccine composition includes MOMP

polypeptide lacking VD1 and VD2, optionally fused to a maltose binding protein. Also claimed are an isolated nucleic acid encoding the polypeptide, a vector, and a method of preventing C. pestis infection by administering the vaccine containing the polypeptide.

CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking  
CC regions VD1 and VD2 are useful for genetic vaccination. The  
CC vaccines are used to prevent *C. psittaci* infection, especially in  
CC birds.

XX  
SQ Sequence 1209 BP; 352 A; 263 C; 260 G; 334 T; 0 other;

. Query Match 47.5%; Score 523.2; DB 20; Length 1209;  
 . Best Local Similarity 69.3%; Pred. No. 1.9e-149;  
 . Matches 790; Conservative 0; Mismatches 308; Indels 42; Gaps



Qy	4	ttcctgttagggaacccctctgtatccaaagcttataataattgatggtacaaatatgggaaggt	63
Db			
Db	67	ttcctgttagggaacccagctgaacccagttattaaatcgatggcactatgtgggaaggt	126
Qy	64	gctgcaggagatccttgagatccttgcgcactcttgctactctgggtgcagcgtattagcttacctgt	123
Db			
Db	127	gcttcaggagatccttgagatccttgccctactctggctactctgggtgagcgcatagcatacccgcca	186
Qy	124	ggatttaccgagagactatgttttcgaccgtatctataaagttagatcacctcaaaacattt	183
Db			
Db	187	ggatactacggagattatgttttcgatcgtgtattaaagttgatgtgaataaaaactttt	246
Qy	184	tctatgggagccaagcctactggatccgctgctgcacactatactact-----	231
Db			
Db	247	agcggcatggctgcaactcctacgcaggctacaggtaacgcaagtaactataatcaagcca	306
Qy	232	---gcgctagatagaaactaaacgcgcctacaaataagaatttacaacgactcagagtaggttc	288
Db			
Db	307	gaagcaaatggcagaaccagcaactcgtctacggaaggaataagggaagatgcagagtggtt	366
Qy	289	actaatgcaggtctcatgtccttaaacatttgggatcgtcttggatttttctctacttta	348
Db			
Db	367	tcaaatgcagccttcctagccttaaacatttgggatcgtctcgacatttactgcacctta	426
Qy	349	ggagcttctaattgttacattagagaaactctacagcgttccaatctcgttggtttattc	408
Db			
Db	427	ggggcatccaaatggatctcaagcaaggttcggctgattccaacttggttgggttaata	486
Qy	409	ggagttaaaggt-----actactgtaaatgcaaatgaactacacaaacggtttct	456
Db			
Db	487	gggttttcagctgaagctcaactctctccgatctccacgcacttccaaacgaactctcaacgagc	546
Qy	457	ttaagtaacggagttgtgaactttacacagacaccttttctcttggagcgtagcgct	516
Db			
Db	547	attaccaaggtgttggaattttatacagacacatcatttctcttggagcgtaggtgca	606
Qy	517	cgtggagccttatgggaatgcggttgtgcaactttgggagctgaattcccaatgcacag	576
Db			
Db	607	cgtggagctttaggaatgggttgcgaacttttagagctgagttcccaatgcagctcaa	666
Qy	577	tccaaacctaaagttagaagaacttaattgtactgtacgtatcgcaattctctgttaaacc	636
Db			
Db	667	tctaactcaagattgaatgtcgaacgtctcaacgtctcaagccagcacaatttggattcac	726
Qy	637	aaaccgaaggcttataaaggcgtt-----gctttcccttggccaaacagacgctggcgta	690
Db			
Db	727	aaaccgaaggcttataaaggcgtagctcgaaatttctcttaccctataacgcttggcaaca	786
Qy	691	gcaacagctactggaacaaagtctgcagccatcaattatcatgaattggcaagtaggagcc	750
Db			
Db	787	acagaagctacagacaccaaatcagctacaaataaatacccatgaatggcaagtaggcct	846
Qy	751	tctctattcaagactaaactcttttagtgcatacatctggagtaacaattggctcagaca	810
Db			
Db	847	gccctgtcttcacagattggaatgtgttccatataatggcgtaaactggtccaagaca	906
Qy	811	actttgatcgtataaacatccgacttgcagccaaactacctaactcagctgtttttaaac	870
Db			
Db	907	actttgatcgtgatactaccgcatctgcacacctaataataaatcggagattcttaac	966
Qy	871	ttaactgcattggaacctcttctactaggaatgcacagcatctgtctactact-----	924
Db			
Db	967	attactacatggaaccaagccttataggtaaccaactgcttggccaataatagtggt	1026
Qy	925	---gattcgtctcagaacttcataagaaattgttctctgcagatacaacaagtttaaatct	981
Db			
Db	1027	aaggatgtctatctgatgtcttgcaattgtcttcgatccagatacaacaaatgaagctct	1086
Qy	982	agaaaacgttggagttactgtaggagctacttttagttgatcgtataaattgctactt	1041
Db			
Db	1087	agaaaacgttgggttagctgtgttggcgaaagcgttaactcgacgtgcacaaatggcgaac	1146
Qy	1042	actgcagaagcctcgttttaataacgagagagcgtcgtcagctatctgttcagttcagattc	1101

Db	1147	actggtgaagcagcgttaataatgaagagctgctcaatgaatgctcaattcagattc	1206
RESULT	4		
AAA64764			
ID	AAA64764	standard; DNA; 1578 BP.	
XX	XX		
AC	AAA64764;		
XX	XX		
DT	02-FEB-2001	(first entry)	
XX	XX		
DE	C. pneumoniae serovar MOMPS pmp gene Ral2 fusion coding sequence.		
XX	XX		
KW	Chlamydial infection; sexually transmitted disease;		
KW	pelvic inflammatory disease; PID; tubal obstruction; infertility;		
KW	trachoma; blindness; acute respiratory tract infection;		
KW	atherosclerosis; coronary heart disease; antibacterial; ss.		
XX	XX		
OS	Chlamydia pneumoniae.		
XX	XX		
PN	WO2000034483-A2.		
XX	XX		
PD	15-JUN-2000.		
XX	XX		
PF	08-DEC-1999;	99WO-US29012.	
XX	XX		
PR	08-DEC-1998;	98US-0208277.	
PR	08-APR-1999;	99US-0288594.	
PR	01-OCT-1999;	99US-0410568.	
PR	22-OCT-1999;	99US-0426571.	
XX	XX		
PA	(CORI-) CORIXA CORP.		
XX	XX		
PI	Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;		
XX	XX		
DR	WPI; 2000-431303/37.		
XX	XX		
PT	Isolated polypeptide for diagnosis and treatment of Chlamydia infection		
PT	comprises immunogenic portion of Chlamydia antigen, which comprises		
PT	amino acid sequence encoded by polynucleotide sequence -		
XX	XX		
PS	Claim 1; Page 205; 256pp; English.		
XX	XX		
CC	The present invention relates to new nucleic acid sequences and the		
CC	proteins encoded by the nucleic acid sequences. The encoded proteins		
CC	comprise an immunogenic portion of a Chlamydia antigen. The encoded		
CC	proteins are useful for the serodiagnosis and treatment of Chlamydia		
CC	infection. Chlamydiae are intracellular bacterial pathogens that are		
CC	responsible for a wide variety of human infections. C. trachomatis		
CC	infection is one of the most common sexually transmitted diseases and		
CC	lead to pelvic inflammatory disease (PID), resulting in tubal obstruction		
CC	and infertility. Trachoma due to ocular infection with C. trachomatis		
CC	the leading cause of preventable blindness worldwide. C. pneumoniae is		
CC	major cause of acute respiratory tract infections in humans and is also		
CC	thought to play a role in the pathogenesis of atherosclerosis and		
CC	coronary heart disease. The present sequence is a nucleic acid sequence		
CC	isolated in the present invention.		
XX	XX		
SQ	Sequence 1578 BP; 414 A; 394 C; 382 G; 388 T; 0 other;		
	Query Match	46.7%;	Score 514.6; DB 21; Length 1578;
	Best Local Similarity	68.8%;	Pred. No. 9e-147;
	Matches 785; Conservative	0; Mismatches 314; Indels 42; Gaps	
QY	3	gttgctgtgagggaaccttctgatccaaagcttataattgattggtgtacataatggaagg	62
Db	435	gctgcctgtgggaatccagctgaaccagtttataatcgatgacctatgtgggaagg	494
QY	63	tgctgcaggagatccttgcgatcctctgcgtactctgcgtactctggtgcagcgtattagattcagtc	122
Db	495	tgtctcaggagatccttgcgatcctctgcgtactctgcgtactctggtgcagcgtattagattcagtc	554



QY 123 tggattttcggagactgttttgcacgctattcttaaaagtagatgcacctaataacatt 182  
 DB 555 aggatactcggagattatgttttcgactgtgtattataaagattgactgtaataaaacttt 614  
 QY 183 ttctatggagcgaacgctactggtatccgctgctgcaaaactataactact----- 231  
 DB 615 tagcgcatggtgcgaactcctacgcaggtctatagtagtaacgcaagtaataactacagcc 674  
 QY 232 ----gctgtagatagacttaacccgctcctacaataagaattttacacgatgcagagtggtt 287  
 DB 675 aagaacaaatggcagaccgaacatcgtctacggaagccatagcaagatgcagagtggtt 734  
 QY 288 cactaatgcaggtctcattgcttaacatttgggactcttctgatttcttctgactt 347  
 DB 735 ttcaaatgcagctcctagccttaaacatttgggactcttctgacatttctgcactt 794  
 QY 348 aggagcttctaattggttaactatagagaactctacagcgttcaactcgttgggtttatt 407  
 DB 795 aggggactcaaatggatactcaaacgaagttcggctgcatcacaacttgggttgaat 854  
 QY 408 cggagttaaaggt-----actactgtaaatgcaaatgaactaccacaaagcttcc 455  
 DB 855 aggggtttcagctgcgaagctcaactctcaccgatcttccaatgcaacttccatacagc 914  
 QY 456 ttaagtaacgaggttggatcttacacagacacacttcttcttctgagcgtagcgc 515  
 DB 915 cattaccgaaggtgtgtggaattttatcacagacacatcatcttcttctgagcgtagcgc 974  
 QY 516 tctgagaccttatggaatgcggtgtgcaacttctggagcgtgaattccaatatgcaca 575  
 DB 975 acgtgagcttctatggaatgtgtgtgcaacttctggagcgtgaattccaatatgcaca 1034  
 QY 576 gtccaacctaagttggaagacttaattgtatctgtacgtatgcgaattctctgttaa 635  
 DB 1035 attcaatcctagattgagatgctcaacgtcacttcaacgcccagcacaatttctgattca 1094  
 QY 636 caaacccaagggtctaaaggcgtt-----gcttcccttgcacacagacgctggcgt 689  
 DB 1095 caaaccaagggtctaaaggcgttagctcgaatttcttccctataacacggtggaac 1154  
 QY 690 acaacacgtactggaacaaatctgcgaccatcaattatcaatgaatgcaagtaggaagc 749  
 DB 1155 acaagagctcacagaccacaaatcagctacaattataatcaccatgaatggcagtaggcct 1214  
 QY 750 ctctctattctacagactaaactctttagtgcatacattgagtagacaatggctcgcagc 809  
 DB 1215 cgcctgtctacagattgaatgtctgttccatataattggcgtaaactggtcgaagc 1274  
 QY 810 aacttttgatgctgataacatccgacttgcagcgaacaaactacacgctgttttaaa 869  
 DB 1275 aacttttgatgctgataacatccgacttgcagcgaacaaactacacgctgttttaaa 1334  
 QY 870 cttaactgcatggaaccccttcttacttaggaatgccaagcattgtctactact----- 924  
 DB 1335 cattactacatggaacccaagccttataggaatacaacactgttttgcacaaataagtgg 1394  
 QY 925 ----gattcgtctcagactcctcaaatgtttctgctcagatacaacaagttaaatc 980  
 DB 1395 taagagttctatcgtatgtcttgcaaatgtcttcgattcagatacaacaagaatgaagtc 1454  
 QY 981 tagaaaagttgtgaggttactgttaggaagcactctttagtgcgtgataaatggtcaact 1040  
 DB 1455 tagaaaagttgtgtgagcgtgtgtgtgcaacgtttaaactgcagcgtgacaaatggtcaat 1514  
 QY 1041 tactgcagagcctcgttttaataacgagagagcgtctcagctatcgttgcagttcagatt 1100  
 DB 1515 cactggtgaagcagcgttaatacaatgaagaagagcgtctcagatgaatgcacaattccgctt 1574  
 QY 1101 c 1101  
 DB 1575 c 1575

RESULT 5  
 AAH56267  
 ID AAH56267 standard; DNA; 1578 BP.  
 XX  
 AC AAH56267;  
 DT 05-SEP-2001 (first entry)  
 XX  
 DE Chlamydia trachomatis pmp gene sequence.  
 XX  
 KW Chlamydia; vaccine; infection; fusion protein; antigen;  
 KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;  
 KW acute respiratory tract infection; Cap1; CT529; OMCB;  
 KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA; ds.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN W0200140474-A2.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PF 04-DEC-2000; 2000WO-US32919.  
 XX  
 PR 03-DEC-1999; 99US-0454684.  
 PR 19-APR-2000; 2000US-0556877.  
 PR 20-JUN-2000; 2000US-0598419.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;  
 XX  
 DR WPI; 2001-374831/39.  
 XX  
 PT Chlamydia polypeptides and fusion proteins useful for preventing pelvic  
 PT inflammatory disease, trachoma, acute respiratory tract infections,  
 PT atherosclerosis and heart disease -  
 XX  
 PS Claim 1; Page 211; 295pp; English.  
 XX  
 CC The present nucleotide sequence is provided in a specification  
 CC relating to compounds and methods for the treatment and diagnosis of  
 CC chlamydial infection. The compounds provided include polypeptides and  
 CC fusion proteins comprising immunogenic portions of Chlamydia antigens  
 CC and DNA sequences encoding such polypeptides. They are useful for  
 CC vaccinating against chlamydial infection, which causes pelvic  
 CC inflammatory disease, trachoma, acute respiratory tract infections,  
 CC atherosclerosis and heart disease.  
 XX  
 SQ Sequence 1578 BP; 414 A; 394 C; 382 G; 388 T; 0 other;  
 Query Match 46.7%; Score 514.6; DB 22; Length 1578;  
 Best Local Similarity 68.8%; Pred. No. 9e-147;  
 Matches 785; Conservative 0; Mismatches 314; Indels 42; Gaps 4;  
 QY 3 gttgctctaggaaccccttctgatacgaacttatttaattggtacataatggaagg 62  
 DB 435 gctgctctggaagaaatcccaagcgaacaaagttatttaacgactatggtggaagg 494  
 QY 63 tctgcagagagatccttgcgactccttgcgactctggtgcgacgctattagcttacgctc 122  
 DB 495 tcttcagagagatccttgcgactccttgcgactctggtgcgacgctattagcttacgctc 554  
 QY 123 tggattttacgagacatgttttgcgacgctatcttaaaagtagatgcacctaataacatt 182  
 DB 555 aggatactcggagattatgttttcgactgtgtattataaagattgactgtaataaaacttt 614  
 QY 183 ttctatggagcgaacgctactggtatccgctgctgcaaaactataactact----- 231  
 DB 615 tagcgcatggtgcgaactcctacgcaggtctatagtagtaacgcaagtaataactacagcc 674  
 QY 232 ----gctgtagatagacttaacccgctcctacaataagaattttacacgatgcagagtggtt 287



```

Db 675 agaagcaaaagcagaccgaacatcgttaccggaaggcatatgcaagatgcagagtggt 734
Qy 288 cactaatgcagcgttcaatgccttaaaacatttggtgagcgttgcgtttctgtacttt 347
Db 735 ttcaaatgcagcgttccatagccttaaaacatttggtgagcgttgcgtttctgcacct 794
Qy 348 agagcgttcaatgccttaacattagaggaacattcagcgttcaatcgttgcgtttatt 407
Db 795 agggcattccaaatgatacttcaaaagcaggttcggtcgtcattcaacttgggtggtta 854
Qy 408 cggagctaaaggt-----actactgtaaatgcaaatgcaactaccacaaacgtttc 455
Db 855 aggtttttcagtcgaagctcaatctctaccgattcttccaaatgcaacttccaaatgag 914
Qy 456 tttaagtaacgaggttggtgaactttacacagacacctttcttcttgagcgttaggcgc 515
Db 915 cattaccacaggttggtggaattttatcacagacacattttcttggagcgttaggtgc 974
Qy 516 tcgtgagcgttatggaatgcgttggtgcaactttggagcgtgaattcccaatatgcaca 575
Db 975 acgtggagcttatgggaatggtggtgcaactttgagcgttaggttcccaatacgcctca 1034
Qy 576 gtccaaacctaagttgaagaacttaattgtagctgtgacgtatgcgaattctctgtaaa 635
Db 1035 attctaactaagattgagatgctcaagctcaacttcaagccagcacaatttgtattca 1094
Qy 636 caaacccaaagggctataaaggcgtt-----gcttcccttggccaaacagcgtggcgt 689
Db 1095 caaaccaagaggctataaaggcgttagctcgaattttcttaccctataacggcgtggaac 1154
Qy 690 agcaacagctactggaacaaagtctgcgaccatcaattatcatgaatggaagtaggagc 749
Db 1155 aacagaagctcacagacacacaaatcagctacaattaaatcaccatgaatggcaagtaggc 1214
Qy 750 ctctctattacagactaaactcttttagtgcatacatattggagtagcaatggctcgagc 809
Db 1215 cgcctgtctacagattgaatgctgtgttccatattgttggttaactggtcaagc 1274
Qy 810 acttttgatgctataacatccgactgctcagcacaacacacacacacacacacacacac 869
Db 1275 acttttgatgctataacatccgactgctcagcacaacacacacacacacacacacacac 1334
Qy 870 cttaactgcatgaaacccctctttactagaaatgcaacagcattgtctactact----- 924
Db 1335 cattactacatgaacccaagccttatagatcaaccactgttggcccaataatgtgg 1394
Qy 925 ----gattcgtctcagacttcatgcaaatgtttctctgcagatacaacaaatgtaaac 980
Db 1395 taaggatgttctatctgctgtcttgcgaattgcttgcagatcaacaaaatgaagtc 1454
Qy 981 tagaaaagcgttggagttactgtgagagcgtactttagttgctgataaaatggtcact 1040
Db 1455 tagaaaagcgttggagttactgtgagagcgtactttagttgctgataaaatggtcact 1514
Qy 1041 tactgagaagcgttgaattaaacagagagcgtcgtcagctatctgttcagttcagatt 1100
Db 1515 cactgtgaagcagcgttaataatgaagagagcgtcgtcagctatctgttcagttcagatt 1574
Qy 1101 c 1101
Db 1575 c 1575

```

RESULT 6

AAV62447

ID AAV62447 standard; DNA; 3133 BP.

XX

AAV62447;

XX

02-FEB-1999 (first entry)

XX

Chlamydia trachomatis major outer membrane protein DNA.

DE

```

XX Major outer membrane protein; MOMP; diagnosis; vaccine; ds.
KW Chlamydia trachomatis serovar L2.
OS
FH Key Location/Qualifiers
CDS 1288..2472
FT sig_peptide /*tag= a
FT mat_peptide /*tag= b
FT /*tag= c
PN US5821055-A.
PD 13-OCT-1998.
XX
XX 06-JUN-1995; 95US-0468451.
XX 13-JAN-1986; 86US-0818523.
XX 14-JAN-1985; 85US-0692001.
XX 25-APR-1991; 91US-0691639.
XX 28-OCT-1993; 93US-0144095.
XX 06-JUN-1995; 95US-0468451.
PA (CHIR ) CHIRON CORP.
PA (WASH-) WASHINGTON RES FOUND.
XX
XX Agabian N, Kuo C, Mullenbach G, Stephens R;
XX WPI; 1998-567652/48.
XX P-PSDB; AAW73141.
XX
XX Claim 5; Fig 2A-E; 15pp; English.
XX
XX This DNA sequence codes for a 42 kDa major outer membrane protein
(MOMP, see AAW73141) of Chlamydia trachomatis serovar L2. A library
of chlamydial genomic DNA was produced in the phage lambda 1059
system. A lambda 1059 recombinant having a 9.2 kb insert was shown
to be homologous to lambda gtl1/L2/33 (see AAV62446) by Southern
analysis and was used for endonuclease digestion mapping and
additional Southern analyses. 2 Contiguous fragments were
identified and these contained sufficient base pairs to encode the
L2 MOMP gene product. These fragments were cloned into M13 for
DNA sequencing. Novel recombinant DNA constructs are provided for
the expression of a polypeptide having immunological activity
corresponding to that of a naturally-occurring MOMP of C.
trachomatis. Such polypeptides find use as reagents in the
detection of C. trachomatis, or antibodies to C. trachomatis, and
as vaccines against infection by C. trachomatis in susceptible
hosts. A claimed probe for detecting C. trachomatis comprises a
polynucleotide fragment that specifically hybridises to a DNA or RNA
sequence encoding C. trachomatis 38-45 kDa MOMP. The probe has
a sequence which is complementary to at least 12 contiguous bases
of the sequence given in AAV62447.
XX
XX Sequence 3133 BP; 911 A; 667 C; 611 G; 944 T; 0 other;

```

Query Match 43.5%; Score 479.4; DB 19; Length 3133;  
Best Local Similarity 66.5%; Pred. No. 6.9e-136;  
Matches 741; Conservative 0; Mismatches 356; Indels 13; Gaps 3;

```

Qy 5 tgccctgagggaaccccttctgacccaagcttattgaatgagtgacaaatagggaggcg 64
Db 1355 tgccctgagggaaccccttctgacccaagcttattgaatgagtgacaaatagggagggt 1414
Qy 65 ctgcaggagatccttgcgataccttgcgtacttgcgtgacgctattagctacgtgcgtg 124
Db 1415 tcggcggagagatccttgcgataccttgcgtacttgcgtgacgctattagctacgtgcgtg 1474

```







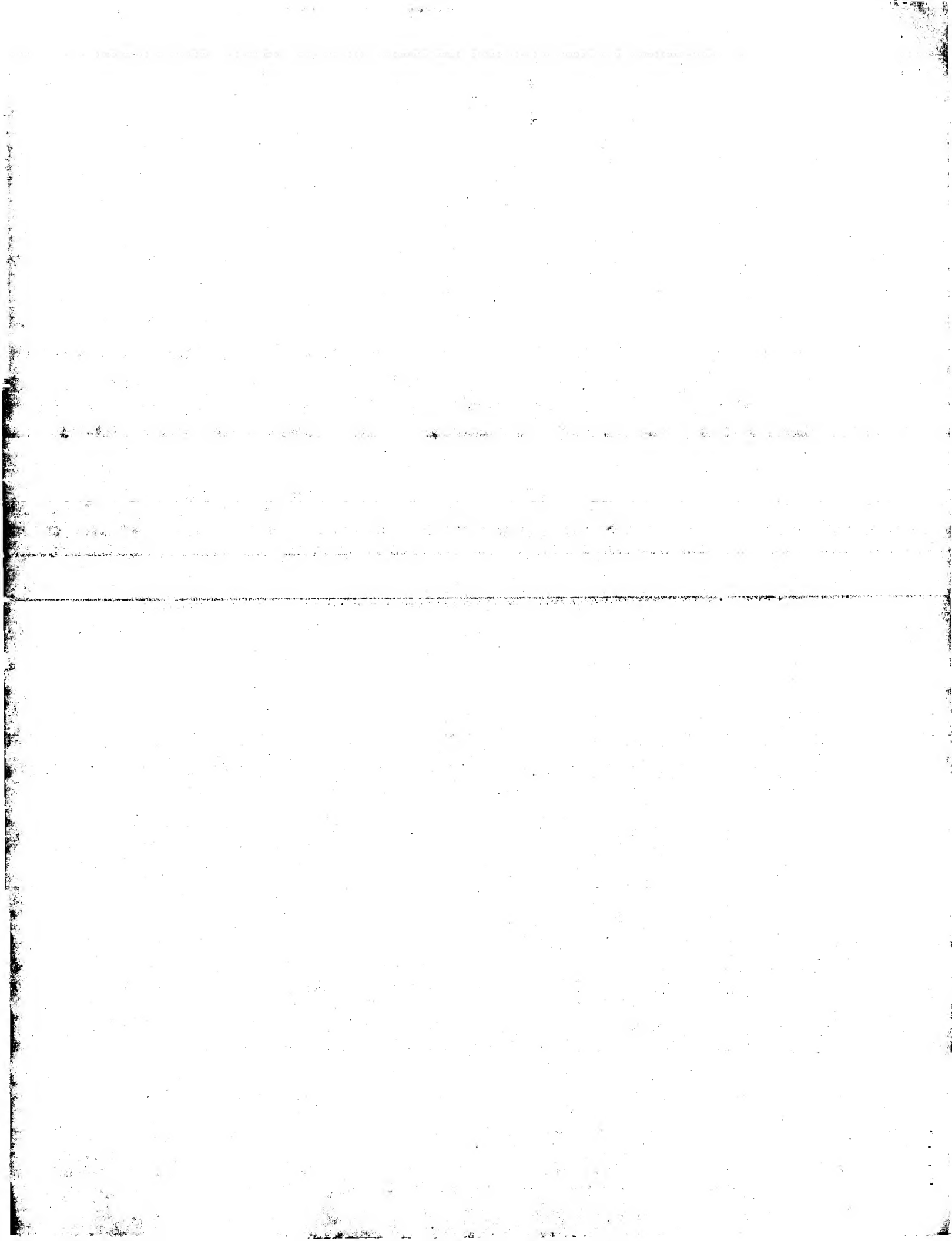
RESULT	8
ID	AAV40646
ID	AAV40646 standard; DNA; 3133 BP.
AC	AAV40646;
XX	
DT	27-OCT-1998 (first entry)
XX	
DE	Chlamydia trachomatis serovar L2 MOMP coding region.
KW	MOMP; major outer membrane protein; immunoassay; diagnosis;
KW	detection; antibody; serovar L2; ds.
OS	Chlamydia trachomatis.
XX	
Key	Location/Qualifiers
CDS	1288..2472
FT	/tag= a
FT	/product= MOMP
FT	1287..1353
FT	/tag= b
XX	
PN	US5770714-A.
XX	
PD	23-JUN-1998.
XX	
PF	14-JAN-1985; 85US-0692001.
XX	
PR	13-JAN-1986; 86US-0818523.
PR	14-JAN-1985; 85US-0692001.
PR	25-APR-1991; 91US-0691639.
PR	28-OCT-1993; 93US-0144095.
PR	06-JUN-1995; 95US-0466814.
XX	
(CHIR )	CHIRON CORP.
PA	(WASH-) WASHINGTON RES FOUND.
PI	Agabian N, Kuo C, Mullenbach G, Stephens R;
XX	
DR	WPL; 1998-376887/32.
DR	p-PSDB; AAW57775.
XX	
PT	DNA coding for Chlamydia trachomatis polypeptide - useful for
PT	producing recombinant polypeptide, etc.
XX	
PS	Claim 5; Fig 2; 15pp; English.
XX	
CC	The sequence is that encoding a major outer membrane
CC	protein (MOMP) of Chlamydia trachomatis. This polypeptide can
CC	be used in immunoassays, e.g. to detect Chlamydia trachomatis
CC	antibodies in blood, or can be used in vaccines. The
CC	polynucleotide can be labelled and used as a diagnostic probe.
XX	
SQ	Sequence 3133 BP; 909 A; 670 C; 610 G; 944 T; 0 other;
	Query Match 43.4%; Score 477.8; DB 19; Length 3133;
	Best Local Similarity 66.4%; Pred. No. 2.1e-135;
	Matches 740; Conservative 0; Mismatches 357; Indels 13; Gaps
Qy	5 tgcctgtagggaacccttctgatcccaagccttaattgattggatgacataatggagaagtg 64
Db	1355 tgcctgtggggaacctgtctgaaaccaagccttatgatcgacggaaattctatgggaagtt 1414
Qy	65 ctgcaggagactcttcgcattcttgccctacttgttgcgcgctattagcttacgtgctg 124
Db	1415 tcggcgagactcttcgcattcttcgcaccttgccacctgtgtgcgcctcatcgatcgctatg 1474
Qy	125 gattttacggagactatgttttcgacccgtatctttaaaagttagatgcacctaaaacatttt 184
Db	1475 gtactatggtgactgttttttcgacccgtgtttttgcaaacagatgtgataaagaattcc 1534



Qy	704	gaacaagtctgcgaccatcaatatcatgaatgcgaagttaggcagctctctattaca	763
Db	971	gaactaaggatgcctctattgattaccatgaatggcaagaagttagctctctcttaca	1030
Qy	764	gactaaactcttagtgccatactatggagtagcaaatgggtctcgagaacttttgatgcctg	823
Db	1031	gactgaatatgttcactccctacattggagtgtaaatgggtctcgagaagttttgatgcag	1090
Qy	824	ataacatccgattgctcagccaaactaccctacagctgttttaaaacttaactgcgatgga	883
Db	1091	acacgattcgtattgctcagccgaagtcagctacaaactgcttttgatgtaccaccttga	1150
Qy	884	accctctttactaggaaaatgccacgcatgtctactactga--ttcgtttctcagact	940
Db	1151	acccaactattgctgagctggcgatgtgaaagctagcgcagaaggtcagctcggagata	1210
Qy	941	tcatgcaaatgttctctgcagatcaacaagtttaaatctagaaaaagcttggagttta	1000
Db	1211	ccatgcaaatcgcttctgcgaatggaacagaagaatctagaaaaactcttgcggtattg	1270
Qy	1001	ctgtaggagctacttttagttgatctgtataaatggctcacttcgcagaagctcgtttaa	1060
Db	1271	cagtaggaaactatttggatgcagacaaatacgcagttacacttgagactcgtttga	1330
Qy	1061	ttaacagagagcgtctcaactatctcgttcggttcagattc	1101
Db	1331	tcgatgagagagctgctcagctaaatgacaaattccgcttc	1371

Search completed: February 7, 2002, 20:23:45  
Job time: 16539 sec







XX A soluble fused protein useful for diagnosis of Chlamydia infection,  
 PT comprises at least part of major outer membrane protein (MOMP) of  
 PT Chlamydia trachomatis ;  
 XX  
 PS Claim 16; Page 28; 37pp; Japanese.  
 XX  
 CC The present invention describes fusion proteins (I) comprising at least  
 CC part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,  
 CC at least one hydrophilic polypeptide having no immunoreactivity to  
 CC human serum and their connected part. AAA08120 to AAA08125 encode  
 CC specifically claimed examples of the fusion proteins given in AY82388  
 CC to AY82393. Also described is a method (A) for the detection of  
 CC Chlamydia trachomatis antibody using (I) as the antigen. (I) is used for  
 CC the diagnosis of Chlamydia trachomatis infectious diseases. The method  
 CC can diagnose Chlamydia trachomatis infectious diseases specifically in a  
 CC high sensitivity.  
 XX  
 SQ Sequence 1362 BP; 383 A; 283 C; 324 G; 372 T; 0 other;

Query Match 26.1%; Score 287.4; DB 21; Length 1362;  
 Best Local Similarity 66.6%; Pred. No. 1.8e-77;  
 Matches 427; Conservative 0; Mismatches 211; Indels 3; Gaps 1;

QY 464 acggaggtgtgaactttacacagacacccctcttcttcttggagcgtgagcgtctggag 523  
 DB 722 acgaaattgtagaaggtatcacagatactactcttcttcttggagcgtgagcgtctggag 781  
 QY 524 ccttatgggaatcggtgtgcaactttggagcgtgaaattccaatgcacagtcacaaac 583  
 DB 782 ctttggggaaatggtgagcgtttaggcgtcttcttcttccaatgcacagtcacaaac 841  
 QY 584 cttaaagtgaagaacttaattgatctgttaacgtatcgcaattctctgttaacaaaccca 643  
 DB 842 cttaaagtgaagaacttaattgatctgttaacgtatcgcaattctctgttaacaaaccca 901  
 QY 644 agggctataaaggcgtgttcttcccttgcacacagcgtggtgagcagcagtcactg 703  
 DB 902 agggatattgaggaagaatccctcttctgtatcttaagcaggaacagatggtgtgacag 961  
 QY 704 gaacaaagtctgcgacacataattcatgaaatggaagcagtaggacgtctctctatcttaca 763  
 DB 962 gaactaagatgcctctattgattaccatgaatggaagcagtaggacgtctctctcttaca 1021  
 QY 764 gactaaactctttagtgcatacatattgagtagcaaatggtctgcgcaacttttgatgctg 823  
 DB 1022 gactgaatatgttcaactcctcatattggagtttaaatggtctgcgcaacttttgatgctg 1081  
 QY 824 ataacatccgcatgtctcagcacaactacacacagcgtgttttaaacattaaactgcagtg 883  
 DB 1082 acacgattcgtattgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1141  
 QY 884 accctctcttactaggaatccacagcagcagcagcagcagcagcagcagcagcagcagcagc 940  
 DB 1142 acccaactattctgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcag 1201  
 QY 941 tcatgcaaatgttctcgtcagatcaacaaagtttaaatctatagaaagcgttgcagtgta 1000  
 DB 1202 ccattgcaaatgttctcgtcagatcaacaaagtttaaatctatagaaagcgttgcagtgta 1261  
 QY 1001 ctgtgagagcactattgattgagcgtgataaaatggtctactactcagaaagcgtcttaa 1060  
 DB 1262 cagttaggaacaaactattgattgagcagacaaatagcaggttactgagactcgttga 1321  
 QY 1061 ttaacagagagcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1101  
 DB 1322 tcgatgagagcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1362

RESULT 15  
 AAA08120  
 ID AAA08120 standard; DNA; 1371 BP.

XX AAA08120;  
 AC  
 XX  
 DT 27-JUN-2000 (first entry)  
 XX  
 DE  
 XX  
 CC trachomatis MOMP containing fusion protein nucleotide sequence #1.  
 KW Chlamydia trachomatis; fusion protein; major outer membrane protein;  
 KW MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;  
 KW infection; infectious disease; ds.  
 XX  
 OS Chlamydia trachomatis.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1371  
 FT /tag= a  
 FT /note= "Fusion protein containing at least part of a  
 FT major outer membrane protein (MOMP) of Chlamydia  
 FT trachomatis; no stop codon given"

JP2000041678-A.

15-FEB-2000.

28-JUL-1998; 98JP-0213212.

28-JUL-1998; 98JP-0213212.

(ELED ) DENKI KAGAKU KOGYO KK.

WPI; 2000-295780/26.

P-PSDB; AAY82388.

A soluble fused protein useful for diagnosis of Chlamydia infection,  
 comprises at least part of major outer membrane protein (MOMP) of  
 Chlamydia trachomatis

Claim 12; Page 26-27; 37pp; Japanese.

The present invention describes fusion proteins (I) comprising at least  
 part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,  
 at least one hydrophilic polypeptide having no immunoreactivity to  
 human serum and their connected part. AAA08120 to AAA08125 encode  
 specifically claimed examples of the fusion proteins given in AAY82388  
 to AAY82393. Also described is a method (A) for the detection of  
 Chlamydia trachomatis antibody using (I) as the antigen. (I) is used for  
 the diagnosis of Chlamydia trachomatis infectious diseases. The method  
 can diagnose Chlamydia trachomatis infectious diseases specifically in a  
 high sensitivity.

Sequence 1371 BP; 390 A; 276 C; 324 G; 381 T; 0 other;

Query Match 26.1%; Score 287.4; DB 21; Length 1371;  
 Best Local Similarity 66.6%; Pred. No. 1.8e-77;  
 Matches 427; Conservative 0; Mismatches 211; Indels 3; Gaps 1;

QY 464 acggaggtgtgaaactttacacagacacccctcttcttcttggagcgtgagcgtctggag 523  
 DB 731 acgaaattgtagaaggtatcacagatactactcttcttggagcgtgagcgtctggag 790  
 QY 524 ccttatgggaatcggtgtgcaactttggagcgtgaaattccaatgcacagtcacaaac 583  
 DB 791 ctttggggaaatggtgagcgtttaggcgtcttcttccaatgcacagtcacaaaccca 850  
 QY 584 cttaaagtgaagaacttaattgatctgttaacgtatcgcaattctctgttaacaaaccca 643  
 DB 851 cttaaagtgaagaacttaattgatctgttaacgtatcgcaattctctgttaacaaaccca 910  
 QY 644 agggctataaaggcgtgttcttcttcccttgcacacagcagcgtggtgagcagcagcagcag 703  
 DB 911 aaggatattgaggcgaagaattccctcttcttgcatttaagcaggaacagatggtgtgacag 970



QY	634	aacaaaccaagcgtctataagaaggctt-----gctttccccccttgccaacagcgctggc	687
Db	184	cacaaaccgaagtactaaagagtagctcgaatttttccttttacctaatacggctggga	243
QY	688	gtagcaaacagctactcgtgaacaaaagctcgcgacctcatattatcatgaatgccagtagga	747
Db	244	acaacagaagctcacagacaccaaatcagctacaatttaaatcacaatgaatggcaagtggc	303
QY	748	gcctctctatttcacagactaaactcttttagtgcatacatattggagtacaaatggctcga	807
Db	304	ctgcgcctgtcttacagattgaatatgcttgttccatatattggcgttaaactggtcaaga	363
QY	808	gcaactttgatcgtgataaacatccgacttgcgcagccaaactacctacagctgttttta	867
Db	364	gcaacttttgatcgtgatactacgcgattgctcaaccttaataataaatcggagattcctt	423
QY	868	aacttaactgcgatgaaccccttttacttagagaaatgccacagcatgtgtctactact---	924
Db	424	aaattactacatgaacccaagccttataggatcaaccactgttigccccataatagt	483
QY	925	-----gattcgtcttcagacttcotatgcaaatgttctcctgtcagatacaacaagttaa	978
Db	484	ggtaagagtgctctatctgatctctgcaaatgtcttcgattcagatccacaacaatgaag	543
QY	979	tctagaaaaagcttggtgagttactgttagagagctacttttagttgatcgtgataaatggt	1038
Db	544	tctagaaaaagcttggtgtagctgttggtgcacacgttaatcgacgctgacaaatggtca	603
QY	1039	cttactgcagaagctcgtttaattaacagagagcgtcgtcacgttatcgttcagttcaga	1098
Db	604	atcacgttgtagaacgcgttaataatgaagagcgtcgcacatgaatgctcaaatcaga	663
QY	1099	ttc t1101	
Db	664	ttc 666	
RESULT 14			
ID	AAA08122	standard; DNA; 1362 BP.	
XX	AA08122;		
AC	AA08122;		
XX	XX	(first entry)	
DT	27-JUN-2000		
DE	C. trachomatis MOMP containing fusion protein nucleotide sequence #3.		
KW	Chlamydia trachomatis; fusion protein; major outer membrane protein;		
KW	MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;		
KW	infection; infectious disease; ds.		
OS	Chlamydia trachomatis.		
OS	Synthetic.		
PH	Key	Location/Qualifiers	
FT	CDS	1...1362	
FT	/**tag= a		
FT	/note= "Fusion protein containing at least part of a		
FT	major outer membrane protein (MOMP) of Chlamy-		
FT	trachomatis; no stop codon given"		
XX	JP2000041678-A.		
PN	15-FEB-2000.		
XX	28-JUL-1998;	98JP-0213212.	
XX	28-JUL-1998;	98JP-0213212.	
PR	(ELED ) DENKI KAGAKU KOGYO KK.		
PA	WPI; 2000-295780/26.		
DR	P-PSDB; AAY82390.		
DR			











Db	1655	cttacatggcaattgaatatttgggactgttttgatgattctgtacattaggaccacca	1714
Qy	359	atggtttacatagaggaactacagoyttcaatctggtttatttcggagttt---	415
Db	1715	gtggatctctaaaggaaatcagcatcttccaacttagttggcttattcggagataatg	1774
Qy	416	-----aagtgactactgttaaatgcaaatgaactaccaaaagtttctttaagtaacggag	469
Db	1775	agaaacctgctacagtttcagatagtaagcttgcatacaaatatgagctctagatcaactg	1834
Qy	470	ttgtgaaactttcacagacaactctttctcttgagagcgttaggcgctctggagccttat	529
Db	1835	ttgttgagttgtatcacagatactacttttgcgtgagtgctggagctctgcagctttgt	1894
Qy	530	ggaaatcggtgtgcacactttggagctgaattccaatatgcacagttccaaactaaag	589
Db	1895	ggaaatggtgagtcgagcttttagcgctcttccaatgcgtccaatccaagctaaag	1954
Qy	590	ttgaagaacttaatgtgactgttaacgtatgcgaattctctgttaacaaacccaagggct	649
Db	1955	tcgaagaattaaacgttctctgtaacgcagctgagtttactatcaataagcctaaagat	2014
Qy	650	ataaaggcgtgttccctctgccaaacagacgctggcgtagcaacagctactggaaaca	709
Db	2015	atgtagggcaagaaatccctctctgtatcttaagcaggacagatgggtgacaggaaact	2074
Qy	710	agtcgcacatcaaatatcatgaatggcaagtggagcgtctctctattcacagactaa	769
Db	2075	aggatgccctattgtattaccataaatggcaagaagtttagctctctcttcacagactga	2134
Qy	770	actctttagtgcaatacatatggagtacaattggctcgagcaacttttgatcgtgataaca	829
Db	2135	atatgttcactccctacattggagttaaatggcttcgagcaagttttgtagcagacaga	2194
Qy	830	tcgcgactgtcagccaaactcacctacagctgttttaaaacttaactgcaggaacctt	889
Db	2195	ttcgtattcgtcagcgagctcagctacaactgtcttttgatgttaccactctgaaccca	2254
Qy	890	ctttactagaaatgcacagcattgtctactactga-----ttcgttcttcagactctatgc	946
Db	2255	ctattcgtgagctgcgatgtgaaagctagcgcagagggctcagctcgagataccatgc	2314
Qy	947	aaattgttctcgttcagatcaacaagttttaaacttagaaaaagctgtgaggattactgtag	1006
Db	2315	aaatcgtttctcgttcgaattgaaacagagtgaaatctagaaaatcttgcggtattgcgtag	2374
Qy	1007	gagctacttttagttgactgtgataaatggcttaacttcgtcagagagctctgttttaatac	1066
Db	2375	gaacaactattgtgagtcagacaaataacgcaggttacaagttgagactcgtcttgatcgtg	2434
Qy	1067	agagagctctcaactatctgtgctagttcagattc	1101
Db	2435	agagagctctcaactgaatgcacattccgcttc	2469

## RESULT

AAZ01425

AAZ01423  
ID AAZ01425 standard; DNA: 1038602 BP.

XX  
XX

AC AAZ01425;

XX  
XX  
-----

DT 07-OCT-1999 (first entry)

XX

DE Complete genome sequence of *Chlamydia trachomatis*.

XX

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis; .....

KW nongonococcal urethritis; epidymitis; cervicitis; salpingitis; .....

KW bartholinitis; pneumop

07-175-175  
XX  
XX

BN	WO9928475-A2.	
XX		
XX	10-JUN-1999.	
ED		
FF		
FF	27-NOV-1998;	98WO-IB01939.
XX		
XX	04-NOV-1998;	98US-0107077.
PR	28-NOV-1997;	97FR-0015041.
PR	17-DEC-1997;	97FR-0016034.
XX		
XX	(GEST ) GENSET.	
PA		
XX		
PI	Griffais R;	
XX		
DR	WPI; 1990-371125/31.	
XX		
XX	Genome sequence of Chlamydia trachomatis	
PT		
XX	Claim 1; Page 373-656; 1755pp; English.	
PS		
XX		
CC	The present sequence represents the complete genome of Chlamydia	
CC	trachomatis. Open reading frames (ORFs) of the genome encode	
CC	polypeptides AA36754-Y37949. The polypeptides can be used as vaccines	
CC	against Chlamydia trachomatis. Antisense and ribozyme sequences can also	
CC	be used to control growth of the microorganism. Chlamydia trachomatis is	
CC	responsible for a large number of diseases, e.g. eye diseases such as	
CC	conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion	
CC	conjunctivitis; genital diseases such as nongonococcal urethritis,	
CC	epidymitis, cervicitis, salpingitis, perihepatitis, Bartholinitis;	
CC	pneumopathy in breast feeding infants; and venereal	
CC	lymphogranulomatosis. The polypeptides of the invention may be of use in	
CC	treating these diseases.	
XX		
XX	Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;	

—



QY	185	ctatgggacgaagccta-----ctgataccgctgctgcaaaactatactactgcgtag	238
Db	1535	aaatgggtgccaagcctcaactgctacaggaagaatgctcagctccatccaattgtaacg	1594
QY	239	atagacctaaocccggcctcaataagaatattacaagctcagagtgagtggttcaactaatgcaag	298
Db	1595	caagagagaatcctgcttacggccgacatagcaggatgctgagatgtttacaatgtctg	1654
QY	299	gcttcattgcttaaaacatttggatcgctttgatgttttctactttttaggagcttcta	358
Db	1655	cttaactggcattgaataatttggatggttttgatgattctctgacattaggagccaaca	1714
QY	359	atggttacattagagaaactctacagcgttcaactcgtttgtttatttcggagttta---	415
Db	1715	gtgatactctaaagaaattcagcatctttcaacttagttggttatttcggagataatg	1774
QY	416	-----aagttactactgtaaatgcaatgaactaccaaaacgtttctttaagtaacggag	469
Db	1775	agaaccatgctacagtttcagatagtaagcttgtaccaaaatagagcttagatcaactcg	1834
QY	470	ttgttaactttacacagacaccttctctctggagcgtaggcgtctgtggagccttat	529
Db	1835	ttgttgatgtatcacagatactacttctgtggatgctggaagctcgcgagctttgt	1894
QY	530	gggaatcggttgtgcaactttgggagctgaattccaatatgcacagtccaaacacctaaag	589
Db	1895	gggaatgtggatgcgcgactttaggcgtctttccaatacgtccaatccaagcctaaag	1954
QY	590	ttgaagaacttaatgtgactgtgaacgtatgcgaattctctgttaacaaacccaaagggct	649
Db	1955	tcaagaattaaaacgttctctgtaacgcagggtgagtttactatacaataagcctaaaggt	2014
QY	650	ataaaggctgtgttcccttgccaacagagcgtggcgtagcaacagctacttggaaaca	709
Db	2015	atgtggggcaagaattccctctctgattctttaagcaggaacaagatggtgtgacaggaaacta	2074
QY	710	agtcgcaacctcaattatcatgaatggcaagtaggagcctctctattctacagactaa	769
Db	2075	aggatgctctattgtattaccatgaatggcaagcaagtttagctctctcttcaagactga	2134
QY	770	actcttagtgcaatacattggagtaacaatggtctcgagcaacttttgatcgtataaca	829
Db	2135	atatgttcaactcctacattggattgaatgaatggctcgagcaagttttgctgagacacga	2194
QY	830	tcgcattgtcgagcaaaactacctacagctgtttttaaacttaactgcatgggaacccctt	889
Db	2195	ttcgtattgtcagcgaagtgcgtacaacgtcttttgatgtttaccactctgaaccca	2254
QY	890	ctttactaggaaatgcacagcaattgtctactactga---ttcgttctcagaacttcatgc	946
Db	2255	ctattgtcgagctggcgatgtgaagcttagcgcaaggggtcagctcgagataccactgc	2314
QY	947	aaattgttctctgcagatacaacaagtttaaatctagaaaagcctgtgtgagttactgtag	1006
Db	2315	aaatcgtttccttgcaattgaaacagatgaatactcagaaaaactcttgcgtttatgcagtag	2374
QY	1007	gagctacttttagttgatgtctgaataatggctacttacttgcagagcctgttttaataacg	1066
Db	2375	gaacaaactattgttgatgcagacaaataacgcagtttaccagttgagactgcgttgatcag	2434
QY	1067	agagagctgtccaagtatcttgggtcagtttcagattc	1101
Db	2435	agadagctgtccacgttaattgcacaattccgcttc	2469

## RESULT 9

RESULTS	5
AAAN60007	
ID	AAAN60007 standard; DNA; 3133 BP.
XX	
XX	
XX	AAAN60007;
XX	
XX	
DT	30-JUL-1991 (first entry)
XX	

DE	Sequence encoding a major outer membrane protein (MOMP) in
L2 B9-F DNA.	
XX	
DE	
XX	
KW	Vaccine; trachoma; conjunctivitis; cervicitis; urethritis; ss.
XX	
OS	Chlamydia trachomatis.
XX	
Key	Location/Qualifiers
FH	1288..1348
CDS	/*tag= a
FT	/product= leader
FT	1349..2472
FT	/*tag= b
FT	misc_difference 2061..2063
FT	/*tag= C
FT	/note= "AGA in AAN60006"
XX	
XX	EP192033-A.
PN	
XX	
PD	27-AUG-1986.
XX	
PD	10-JAN-1986; 86EP-0100279.
PF	
XX	
PR	14-JAN-1985; 85US-0692001.
XX	
XX	(CHIR- ) CHIRON CORP.
PA	
PI	Agabian N, Stephens R, Kuo CC, Mullenbach GT;
XX	
DR	WPI: 1986-226702/35.
DR	P-FSDB; AAF60004.
XX	
XX	New DNA constructs and polypeptide(s) - displaying antigenicity
PT	of major outer membrane protein of Chlamydia trachomatis
PT	
PT	
XX	
PS	Example; Appendix B; 31pp; English.
XX	
CC	DNA obtd. from C.trachomatis serovar L2 was partially digested with
CC	DNAse I and inserted into vector lambda gtl1. Clone lambda gt11/L2/
CC	33 (AAN60006) reacted with the pool and was subsequently shown to
CC	produce a polypeptide that displays species-, subspecies- and type-
CC	specific epitopes of the chlamydial MOMP. L2 B9-F DNA (AAN60007) is
CC	comprised of lambda 1059 recombinants shown to be homologous with
CC	lambda gtl1/L2/33.
XX	
XX	Sequence 3133 BP: 911 A; 673 C; 606 G; 943 T; 0 other;

Query Match	43.1%;	Score	474.6;	DB	7;	Length	3133;
Best Local Similarity	66.2%;	Pred.	No. 2e-134;				
Matches	738;	Conservative	0;	Mismatches	359;	Indels	18;
Gaps	3						
Qy	5	tgcctgtagggaaaccccttctgatccaaagccttatttaattgatgggtacaatatgggaaggtg	64				
Db	1355	tgcctgtgggaatcctgtctgaaccaagccttatgatcgacggaaattctctatgggaeggtt	1414				
Qy	65	ctgcagagatccttcgcattcttcgctacttgccttgcgcacgctatttagcttgcgtgctg	124				
Db	1415	tcgscgacatccttcgcattcttcgaccttgcaccttgcgtgcgcgtatcaagatgcgtatgg	1474				
Qy	125	gattttacggagactatgttttcgacgctatctctaaagtagatgcacctataaacatttt	184				
Db	1475	gtactatggtgacttgttttcgacccgtgttttgcaaacagatgtgaataaagaattcc	1534				
Qy	185	ctatggagcgaagccta-----ctggatcgcgtctgtgcaaacctatactactgccctag	238				
Db	1535	aaatgggtgccaagcctcacactgtctacaggccaatgctgcagctccatccactgtacag	1594				
Qy	239	atacacctaaccggcctcaataaagcatttacacagatcgagatgggttcactaatgcag	298				
Db	1595	caagagagaatcctgtcttaccgcccgcacatatgcaggatgctgagatgittacaaatgctg	1654				
Qy	299	gcttcattgccttaaacatttggatcgcttcttgatgttttctgtacttttaggagcttcta	358				



Query Match	99.7%	Score 1098	DB 1	Length 1830
Best Local Similarity	100.0%	Pred. No. 1.1e-275		
Matches 1098	Conservative	0	Mismatches	0
			Indels	0
				Gaps
QY	4	ttgcctgtaggaacccctctgacccaagccttattaatgatggtacaatatggaaggt	63	
DB	386	TTGCCTGTAGGAACCCCTCTGATCCAAAGCTTATTATTGATGGTACAAATATGGAAGGT	445	
QY	64	gctcaggaagatccttgcgataccttgcgtacttggctgcagcgtattagcttcagtgct	123	
DB	446	GCTCAGGAGATCCTTCGATCTCTGGGCTACTTGGTGGCAGCGTATTAGCTTACGTGCT	505	
QY	124	ggattttacggagactatgttttcgaccgtatctctaaaagttagatgcactaaacattt	183	
DB	506	GGATTTTACGGAGACTATGTTTCGACCGTATCTTAAAGTAGATGCACCTAAACATTT	565	
QY	184	tctaaggagccaaagcctactggtatccgctgctgcgcaactactactgcgtagataga	243	
DB	566	TCTATGGAGCCAAAGCCTACTGGATCGCGCTGCTGCAAACTATACTACTGCGTAGATAGA	625	
QY	244	cctaaccggcctacaataaagcatttacacgagtcagagtggttccactaatcagagcttc	303	
DB	626	CCTAAACCGGCCCTACAAATAGCATTTACAGATGCAGATGGTTCACATATGACGGCTTC	685	
QY	304	attgccttaaacattttgggatacgtcttgatgtttcttctgactttgaggagcttctaaggt	363	
DB	686	ATTGCCCTTAAACAATTTGGGATCGCTTGGATGTTTCTGTACTTTTGGAGAGCTTCTAATGGT	745	
QY	364	tacattagagaaactctacagqgttcaactcgttggtttattccggagttaaaggtact	423	
DB	746	TACATFAGAGAAACTCTACAGCGTTCAATCTCGTTGCTTTATTCGGAGTTAAAGGTACT	805	
QY	424	actgtaaatgcgaatgaactaccaaacytttctttaaagtaacggagttgttgaactttac	483	
DB	806	ACTGTAATGCMAATGAACATACCAACAGCTTCTTTAAGTAACGGAGTGTGTGAACCTTAC	865	
QY	484	acagacacctcttctcttggagcgtagcgctgtaggccttattaggaatcggttgt	543	
DB	866	ACAGACACCTCTTCTCTTGGAGGTAGGCGCTCTGTCGGACCTTATGCGGAATGCGGTGT	925	
QY	544	gcaactttggagcgtgaattccaatgacagtcacaaacctaaaagttgaagaacttaat	603	
DB	926	GCAACTTTGGGAGCTGAATTCCAATATGCAGTCCAAACCTTAAAGTTGAAGAACTTAAAT	985	
QY	604	gtgatctgaacgtatcgcaattctctgtaaacaaaccccaaggcgtataaagcgttgtct	663	
DB	986	GTGATCTGAACGTATCGCAATCTCTCTTAACNAACCCAAAGGCGTATAAAGCGTTGCT	1045	
QY	664	ttcccttgcacaacagcgtgcgtgtagcaacagctactggaacaaagctcgcgacatc	723	
DB	1046	TTCCCTTGGCAACAGACGCTGGCGTAGCAACAGCTACTTGGAAACAAAGTCTGGACCATC	1105	
QY	724	aattatcatgaatggcgaagtagagcctctctattcacagactaaactcttagtgcca	783	
DB	1106	AATTATCATGNAAGGCAAGTAGAGCCCTCTCTCTTCTTACAGACTTAAACTCTTTAGTGCCA	1165	
QY	784	tacattggaggtacaatgggtctcgcagcaacttttgatgctgataacaatcccgattgctcag	843	



12]	Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA	Takeuchi A., Nishida J., Shibata K., Fujinaga R., Itoneda H., Matsushima H.,
RA	Hanaka C., Furukawa S., Miura K., Nakazawa A., Ishii K., Shiba T.,
RA	Hattori M., Kuhara S.;
RT	"Comparison of outer membrane protein genes omp and pmp in the whole genome
RT	sequences of Chlamydia pneumoniae isolates from Japan and US";
RL	Unpublished.
XX	
XX	SWISS-PROT; P27455; OMP1_CHLPN.
XX	
XX	key Location/Qualifiers
PH	
PH	
FT	1. .1170
FT	/db_xref="taxon:83558"
FT	/note="synonym:Chlamydia pneumoniae"
FT	/sequenced_mol="DNA"
FT	/organism="Chlamydia pneumoniae"
FT	/strain="J138"
FT	1. .1170
FT	/codon_start=1
FT	/db_xref="SWISS-PROT:P27455"
FT	/transl_table=11
FT	/gene="ompA"
FT	/product="major outer membrane protein"
FT	/protein_id="BAA85940.1"
FT	/translation="MKKILKSALLSAAFGSVGSLQALFVGNPSPDPSLLIDGTIWEGAA
FT	GPDCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTFSGMAKPTGSAANYTTAVDFR
FT	NPAYNKLHDAEWFNTNAGFIALNIWDRDVFCTLGASNGYIRGNTAFNLVLGFVGKT
FT	TYNANLPHVSLISNGVELYDTSPFSWSVGARGALWEGCGCATLGAEFOYAGASQKPKVEL
FT	NVICNVQSQSNKPKGYGVAEPPLPDTAGVATATGKATINHEWFOYASISYRLNLS
FT	VPYIGVQWSRATFDANIRIAQPLKPLTAVLNLTANNPILLGNATLSITDPSFDMQIV
FT	SCQINKFKRKACGVTVGATLVDAKWSLTAEARLNIERAAHVSGQFR"
XX	
XX	Sequence 1170 BP: 320 A; 255 C; 246 G; 349 T; 0 other;
XX	

	Query Match	99.7%; Score 1098; DB 15; Length 1170;	
	Best Local Similarity	100.0%; Pred. No. 1.le-275;	
	Matches 1098; Conservative	0; Mismatches	0; Gaps
		Indels	0;
Qy	4	t t g c g t g a g g a a c c c t t g a c c a a g g t t a t t a a t g a t g t a c a a t a t g g g a a g g t	63
Dd	70	T T G C C T G A G G A A C C C T C T G A T C C A A G C T T A T T A A T T G A T G G T A C A A T A T G G A A G G T	129
Qy	64	.g t c g a g g a g a c o b t t g c g a t c o t t i g c g t a c t t g t g c g a c o c t a t t a g c t t a c t g e t	123
Dd	130	G C T G C A G G A G A T C C T T G C G A T C C T T G G C G T A C T T G G T G C G A C C T A T T A G C T T A C G T G C T	189
Qy	124	g g a t t t a c g g a g a c t a t g t t t c g a c c g t a t c t c t a a a a g t a g a t g c a c c t a a a c a t t t	183
Dd	190	G G A T T T T A C G A G A C T A T G T T T C G A C C G T A T C T T A A A G T A G A T G C A C C T A A A A C A T T T	249
Qy	184	t o t a t g g g a g c c a a g c o c t a c t g g a t c o c y t g t g c a a a c t a c t a c t g c g t a g a t a g a	243
Dd	250	T C T A T G G G A G C A A G C C T A C T G S A T C G C G T G C T G C A A A C T A T A C T A C T G C G G T A G A T A G A	309
Qy	244	c t a a c c c g c c t a a a t a a g c a t t t a c a g a t g c a g a t g g t t c a c t a a t a c g a g g t t c	303
Dd	310	C C T A A C C C G G C C T A C A A T A A G C A T T A C A G A T C A G A G T G G T C A C T A A T G C A G G C T T C	369
Qy	304	a t t g o c t t a a a c a a t t g g g a t g e g t t t g a t g t t t c t g t a c t t t a g a g a g o c t c a a t g g t	363
Dd	370	A T T G C C T T A A A C A T T T G G G A T C G C T T T G A T G T T T C T A C T A T T A G G A G C T T C T A A T F G G T	429
Qy	364	t a c a t t a g a a a a c t c f a c a g c g t c a a c t c t g t t g t t a t t c g a g t t a a a g g t a c t	423
Dd	430	T A C A T T A G A G G A A A C T C A C A C G T T C A A T C C G T T G G T T A T T C G G A G T T A A A G G T A C T	489
Qy	424	a c t g t a a a t g c a a a t g a a c t a c c a a a o g t t c t t a a g t a a c g g a g t g t t g a a c t t a c	483
Dd	490	A C T G T A A A T G C A A A T G A A C T A C C A A A C C G T T C T T T A G T A A C G A G A G T G T T G A A C T T T A C	549

QY	484	acagacacacctttcttcttgaggcgttaggcgctcgtggagcccttatgggaatcggggttgt	543
Db	550	ACAGACACACTCTTTCTCTGGAGCGTAGCGCCTGCTGAGCCCTATGGGAATCGCGTTGT	609
QY	544	gcaactttggggagctgaattccaatatcacagtccaaacctaaaagttgaagaactaat	603
Db	610	GCAAACTTTGGGAGCTGAATTC AATATGCACAGTCCAACCCTAAGTTGAAGAACAATTAAAT	669
QY	604	qtgactgttaacgcatcgcaattctctgtataacaaccacaaaggcgtataaagcgcttgct	663
Db	670	GTGATCTGTAAACGTATCGCAATCTCTGTAAACAACCCAGGCGCTATAAAGCGCTTGCT	729
QY	664	tcccccttgcacaacagcgtgcgtagcaacagctactctggaacaaagtcctgeaacatc	723
Db	730	TTCCCCTTGCCAACAGACGCTGGCGTAGCAACAGCTACTGGAACAAGTCTGGACCATC	789
QY	724	aattatcataatgccaadtaggagccctctctatttacagactaaactottitaggcca	783
Db	790	AATTATCATGAATGGCAAGTAGGAGCCCTCTATCTTACAGACTAAACTCTTTAGTGCCA	849
QY	784	tacatgtgagtaacaatgctctcgagcaacttttgaatcgtgataacatcccagctgccag	843
Db	850	TACATGGAGTACAATGGTCTCGAGCAACTTTTGATGCTGATACATCCGCATTCCTCAG	909
QY	844	ccaaaactactacagcgtgtttaaaacttaaactgatggaacccctcttactaggaaat	903
Db	910	CCAAACTACTACAGCTGTTTTAAACTTAACCTGCATGGAACCCCTCTCTTACTAGGAAAT	969
QY	904	gccacagcattgtctactactgattogttctcagactctcatgcaaatgtttcctgtcag	963
Db	970	GCCACAGCATTTGCTACTACTGATTCGTTCTCAGAACTTCATGCAAAATGTTTCTCTGT	1029
QY	964	atcaacaagtttaaatctagaaaaagctgtgaggtactctagagctcactttagttgat	1023
Db	1030	ATCAACAAGTTTAAATCTAGAAAGCTTTGGAGATTACTGTAGGAGCTACTTTAGTTGAT	1089
QY	1024	gctgataatggctcacttactcagagagctcgttttaattaacgagagagctgcacgta	1083
Db	1090	CCTGATAAATGGTCATCTACTGCAAGAGCTCGTTTAAATTACGAGAGAGCTGCTCACGTA	1149
QY	1084	tctggtcagtttcagattc 1101	
Db	1150	TCTGGTCAGTTCAGATTC 1167	
RESULT	3		
CHTMOMPP			
LOCUS	CHTMOMPP	1830 bp	DNA
DEFINITION	Chlamydia pneumoniae major outer membrane protein precursor,	BCT	26-APR-1993
complete cds.			
ACCESSION	M69230		
VERSION	M69230.1	GI:144540	
KEYWORDS	major outer membrane protein.		
SOURCE	Chlamydia pneumoniae DNA.		
ORGANISM	Chlamydophila pneumoniae		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.		
AUTHORS	1 (bases 1 to 1830)		
TITLE	Perez Melgosa, M., Kuo, C.-C. and Campbell, L.A.		
FEATURES	Sequence analysis of the major outer membrane protein gene of		
JOURNAL	Chlamydia pneumoniae		
MEDLINE	Infect. Immun. 59, 2195-2199 (1991)		
FEATURES	91244474		
source	Location/Qualifiers		
	1..1830		
	/organism="Chlamydophila pneumoniae"		
	/db_xref="taxon:83558"		
RBS	306..310		
	/note="putative"		
sig_peptide	317..385		
CDS	317..11486		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA73071.1"		



JOURNAL Submitted (26-FEB-1999) Pathology, Vanderbilt University, C-3321  
 Medical Center North, Nashville, TN 37232, USA

FEATURES  
 Location/Qualifiers  
 Source  
 1..1170  
 /organism="Chlamydomonas reinhardtii"  
 /db\_xref="taxon:83558"  
 /note="from the cerebral spinal fluid of a patient with rapidly progressive multiple sclerosis"  
 1..1170  
 /gene="OMP"  
 /gene="OMP"  
 1..1170  
 /gene="OMP"  
 /codon\_start=1  
 /transl\_table=11  
 /product="major outer membrane protein"  
 /protein\_id="AAD22492.1"  
 /db\_xref="GI:4545321"  
 /translation="MKLLKSLLSAFAFSGVSLQALPVGNPDSPDLLIDGTWEGA  
 AGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDAPKTFSGAKPTGSAANVYTTAVD  
 RPNPAYNKHLDHAEFTNAGFIALNINDRFDVFTLGASNGYIRGNSTAFNLVLFV  
 KGTVTNANLHNPVSLNGVELYDTSPFSWVGARGALWECGATLGAEFOYAOAKPK  
 VEELNVICNVSOFSVNGPKYGVAFPLPTDAGVATGTGKATINHEWVGASLSY  
 RLNSLVPYIGVQWRSRATFADNIRLQKPLTAVLNLTAHNPSSLGNATLSTDSRS  
 DPMQIVSQINKFKSRKACGVTVGATLVADKWSLTAERLINERAAHVSGQFRP"  
 54  
 variation  
 /gene="OMP"  
 /note="silent mutation"  
 /replace="c"  
 126  
 variation  
 /gene="OMP"  
 /note="silent mutation"  
 /replace="g"  
 350 t

BASE COUNT 320 a 254 c 246 g 350 t

ORIGIN

Query Match 99.7%; Score 1098; DB 1; Length 1170;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-275;  
 Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttgctgttagggagcccttctgacccaagcttataatgattggtgacataatgggaagt 63  
 DB 70 TTGCTGTAGGAGAACCTTCTCATCAAGCTTATTAAATGATGGGTACAAATATGGAGGT 129  
 QY 64 gctgagagagatccttgagatccttgcctacttgcctgctgagcgcctattagcttgcgt 123  
 DB 130 GCTGAGGAGATCCTTGGCGATCCTTGGCTACTTGGTGGCGAGCTATTAGCTTACGTGT 189  
 QY 124 ggattttacggagactatgtttgacccgtatcttataaagtagatgcacataaacattt 183  
 DB 190 GGATTTACGGAGACTAATGTTTCGACCGTATCTTAAAGTAGATGCACCTAAACATTT 249  
 QY 184 tctatgggagccaaactactgagctgctgctgcaactataactactgctgagtagataga 243  
 DB 250 TCTATGGGAGCCAAAGCTTACTGGATCGCTGCTGCAAACTACTACTTGGCGTAGATAGA 309  
 QY 244 cctaaccggcctacaataagcattacacgattgacagagtggttcacttaagcaggttc 303  
 DB 310 CCTAACCCGGCCTACAATAAGCATTTACAGATGACAGATGCTGTTTCAATATGACGGCTTC 369  
 QY 304 attgcttaaacatttggagatccttgcctgctgctgctgctgctgctgctgctgctgct 363  
 DB 370 ATTGCTTAAACATTTGGGATCGCTTGTGATGTTTCTGCTGCTGCTGCTGCTGCTGCT 429  
 QY 364 tacattaggaagaaactacagcttcaatcctgctgctgctgctgctgctgctgctgctgct 423  
 DB 430 TACATTAGAGAACTCTACAGCTTCAATCTCGTTGGTTTATTTCGGAGTTAAAGGTACT 489  
 QY 424 actgtaaatgcaatgaactacaaacgcttctttaaagtaacgaggtgtgtaaacattac 483  
 DB 490 ACTGTAAATGCAATGAACACTCAACACGTTCTTTTAAAGTAACGGAGTGTGTTGAACCTTAC 549  
 QY 484 acagacacacctcttctcttctgagcgtagcgcgtctgagccttatgggaatgcggttgt 543

DB 550 ACAGACACTCTTTCTCTTGGAGCGTAGCGCTCTGTGGAGCTTATGGAAATGCGGTGT 609  
 QY 544 gcaacttgggagctgaattccaatatgcacagtcaccaacacctaagttgaagaactaat 603  
 DB 610 GCAACTTTGGGAGCTGAATTCCTCAATATGCACAGTCCAAACCTTAAAGTTGAAGAATAAT 669  
 QY 604 gfgactctgaactgacgaattctcttaacaaacccaagggcgtataaaggcgttgtct 663  
 DB 670 GTGATCTGTACGTATCGCAATTCCTGTAAACAAACCAAGGGGCTATAAAGGCGCTTGT 729  
 QY 664 ttcccttggccaacagacgctggtgctgacacagctactggaacaaaagtcgacaccatc 723  
 DB 730 TTCCCTTGGCCAACAGACGCTGGCGTAGCAACAGACTTGGAAACAAAGTCTGCGACCATC 789  
 QY 724 aattatcatgaatggcgaagtaggagcctctcttatctacagactaaaactcttttagtcca 783  
 DB 790 AATTATCATGAATGTCGAAGTAGGAGCCTCTCTATCTTACAGACTAAACTCTTTTAGTGCA 849  
 QY 784 tacattggagacaaatgctcgcagacaaacttttgcctgataacacatcgcaattgctcag 843  
 DB 850 TACATTGGAGTACAAATGGTCTCGACCAACTTTTGTATGCTGATAACATCCGCAATTGCTCAG 909  
 QY 844 ccaaaactacctaagcgtgttttaaaacttaactgcatggaacccctcttttactaggaat 903  
 DB 910 CCAAAACTACCTACAGCTGTTTAAACTTAACCTGCATGGAACTCTTTTACTAGGAAT 969  
 QY 904 gccacagatgctactactgattcgttctcagactcagactcagactcagactcagactcag 963  
 DB 970 GCCCAGCATTTGCTTACTTACTTCTGATTCGTTCTCAGACTTTCATGCAAAATTTGTTCTCTCAG 1029  
 QY 964 atcaacaagtttaaaactagaaaagcttgcgtgagttactgtaggagctactttagttag 1023  
 DB 1030 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT 1089  
 QY 1024 gctgataaaggctcacttactgacgagaagctcgttttaataacagagagcgtcgtcacgta 1083  
 DB 1090 GCTGATAAATGGTCACTTACTTCTCAGAGAGCTCGTTTAAATTAACGAGAGAGCTGCTCACGTA 1149  
 QY 1084 tctgctcagttcagattc 1101  
 DB 1150 TCTGCTCAGTTCAAGATTC 1167

RESULT 2  
 AB033787  
 ID AB033787 standard; DNA; PRO; 1170 BP.  
 AC AB033787;  
 SV AB033787.1  
 DT 04-NOV-1999 (Rel. 61, Created)  
 DT 04-NOV-1999 (Rel. 61, Last updated, Version 1)  
 DE Chlamydomonas reinhardtii ompA gene for major outer membrane protein,  
 DE complete cds.  
 XX major outer membrane protein.  
 XX Chlamydomonas reinhardtii  
 OS Chlamydomonas reinhardtii  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomonadales.  
 RN 1-1170  
 RP Hirakawa H., Shirai M.;  
 RT Submitted (25-OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RL Hideki Hirakawa, Kyushu University, Graduate School of Genetic Resources  
 RL Technology; 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan  
 RL (E-mail:hirakawa@grt.kyushu-u.ac.jp, URL:http://www.grt.kyushu-u.ac.jp,  
 RL Tel:81-92-642-3043, Fax:81-92-642-3043)  
 XX



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:37:19 ; Search time 12230.8 Seconds  
(without alignments)  
1485.051 Million cell updates/sec

Title: US-09-391-606-14  
Perfect score: 1101  
Sequence: 1 agtgcctgttaggaaccc.....tatctggtcagtcagattc 1101

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.om.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.sy.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htgo.hum.\*
- 31: em.htgo.inv.\*
- 32: em.htgo.rod.\*
- 33: em.htg.hum.\*
- 34: em.htg.inv.\*
- 35: em.htg.rod.\*
- 36: em.htg.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1098	99.7	1170	1	AF131889	AF131889 Chlamydia
2	1098	99.7	1170	15	AB033787	AB033787 Chlamydia
3	1098	99.7	1830	1	CHTOMOPP	M69230 Chlamydia p
4	1098	99.7	20057	1	AE001652	AE001652 Chlamydia
5	1098	99.7	25150	1	AE002167	AE002167 Chlamydia
6	1098	99.7	300550	1	AP002547	AP002547 Chlamydia
7	1096.4	99.6	1170	1	CHTOMPB	M64064 Chlamydia p
8	1086	96.8	1602	1	CPMOMPX	X72023 Chlamydia p
9	997.2	90.6	1170	1	CHTOMPEQ	L04982 Chlamydia p
10	967	87.8	999	1	CHTOMPAAI	M73038 Chlamydia p
11	668.4	60.4	670	1	AF131230	AF131230 Chlamydia
12	665.2	60.4	670	1	AF131229	AF131229 Chlamydia
13	606.6	55.1	1425	1	AF269259	AF269259 Chlamydia
14	600.2	54.5	1170	1	AF272945	AF272945 Chlamydia
15	600.2	54.5	1261	1	CHTOMPAAD	M73036 Chlamydia p
16	600.2	54.5	1338	1	CPOMP	X51859 Chlamydia p
17	600.2	54.5	1338	1	CHTOMPAAAA	L39020 Chlamydia p
18	600.2	54.5	1435	1	AF269256	AF269256 Chlamydia
19	595.8	54.1	1436	1	AF269267	AF269267 Chlamydia
20	592.6	53.8	1430	1	AF269266	AF269266 Chlamydia
21	590.6	53.6	1392	1	AF269279	AF269279 Chlamydia
22	585.8	53.2	1176	1	CHTOMPEXA	L25436 Chlamydia p
23	585.8	53.2	1412	1	AF269260	AF269260 Chlamydia
24	585.8	53.2	1421	1	AF269261	AF269261 Chlamydia
25	581.2	52.8	1482	1	AF269282	AF269282 Chlamydia
26	573.4	52.1	1200	1	AF269257	AF269257 Chlamydia
27	573.4	52.1	1200	1	AF269258	AF269258 Chlamydia
28	573.4	52.1	1227	1	CHTOMPAAH	M73037 Chlamydia p
29	573	52.0	1387	1	AF269280	AF269280 Chlamydia
30	571.8	51.9	1343	1	CPFFNMOMP	X61096 Chlamydia p
31	555.6	50.5	1434	1	AF269269	AF269269 Chlamydia
32	552.2	50.2	1107	1	CPDNAX	Z18756 Chlamydia p
33	543	49.3	1024	1	CHTOMPAAK	M73040 Chlamydia p
34	536	48.7	1058	1	CPAJ5617	AJ005617 Chlamydia
35	534.4	48.5	1058	1	CPAJ5613	AJ005613 Chlamydia
36	534.4	48.5	1058	1	CPAJ5614	AJ005614 Chlamydia
37	532.8	48.4	1058	1	CPAJ5615	AJ005615 Chlamydia
38	532.8	48.4	1058	1	CPAJ5618	AJ005618 Chlamydia
39	532.8	48.4	1058	1	CPJ004875	AJ004875 Chlamydia
40	529.6	48.1	1445	1	AF269265	AF269265 Chlamydia
41	529.6	48.1	1610	1	CHTOMPPXX	L04980 Chlamydia p
42	528	48.0	1058	1	CPJ004874	AJ004874 Chlamydia
43	528	48.0	1393	1	AF269262	AF269262 Chlamydia
44	528	48.0	1432	1	AF269264	AF269264 Chlamydia
45	528	48.0	1438	1	AF269263	AF269263 Chlamydia

## ALIGNMENTS

### RESULT 1

AF131889	AF131889	1170 bp	DNA	BCT	31-MAR-1999
LOCUS	Chlamydia pneumoniae major outer membrane protein (MOMP) gene,				
DEFINITION	complete cds.				
ACCESSION	AF131889				
VERSION	AF131889.1	GI:4545320			
KEYWORDS	Chlamydia pneumoniae.				
SOURCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia p.				
ORGANISM	Chlamydia pneumoniae.				
REFERENCE	1 (bases 1 to 1170)				
AUTHORS	Sriram,S., Mitchell,W.M. and Stratton,C.W.				
TITLE	Multiple sclerosis associated with Chlamydia pneumoniae infection of the CNS				
JOURNAL	Neurology 50 (2), 571-572 (1998)				
MEDLINE	98145402				
REFERENCE	2 (bases 1 to 1170)				
AUTHORS	Mitchell,W.M., Tharp,A.C., Stratton,C.W. and Sriram,S.				
TITLE	Direct Submission				



GenCore: version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:38:07 ; Search time 96.2 Seconds  
(without alignments)  
366.620 Million cell updates/sec

Title: US-09-391-606-9  
Perfect score: 2261  
Sequence: 1 MWNPIGPGIDETERTPPAD.....SAAVVSAGVLPLOQVLWIRA 463

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2214	97.9	715	2 I40729	hypothetical 76k p
2	2188	96.8	651	2 D72042	conserved hypothet
3	2188	96.8	651	2 E86581	CHLPN 76 kDa homol
4	451.5	20.0	647	2 G71490	hypothetical prote
5	201	8.9	973	2 C85693	probable membrane
6	188	8.3	1122	2 G64887	probable tail fibe
7	182.5	8.1	2055	2 T31110	extracellular matr
8	177.5	7.9	1822	2 S33441	EF protein - Strep
9	174.5	7.7	545	2 E84327	Htr7 transducer [i
10	174.5	7.7	545	2 T46811	halobacterial tran
11	171	7.6	627	2 F84194	Htr14 transducer [
12	166.5	7.4	1128	2 T30296	R27-2 protein - Tr
13	166.5	7.4	2232	2 T34434	hypothetical prote
14	160.5	7.1	544	2 T44938	transducer protein
15	158.5	7.0	1156	2 T34852	probable secreted
16	158	7.0	536	1 A47190	transducer protein
17	158	7.0	536	2 E84318	Htr1 transducer [i
18	158	7.0	810	2 T46810	halobacterial tran
19	158	7.0	810	2 F84327	Htr5 transducer [i
20	158	7.0	1365	2 T30822	lmp1 protein - Myc
21	156.5	6.9	641	2 C82206	methyl-accepting c
22	156	6.9	990	2 I51618	nucleolar phosphop
23	155	6.9	892	2 T50985	related to transcr
24	154.5	6.8	764	2 A84328	Htr2 transducer [i
25	154.5	6.8	765	1 T44946	transducer protein
26	154.5	6.8	1262	2 T22523	hypothetical prote
27	152.5	6.7	860	2 T14650	tail fiber protein
28	152.5	6.7	978	2 T14968	phage lambda-relat
29	152	6.7	1561	1 S06839	surface antigen sp

30 151.5 6.7 582 2 S24545 intermediate filam  
31 151.5 6.7 1238 2 T03465 probable exonuclea  
32 150.5 6.7 1528 2 A60338 surface antigen A  
33 149 6.6 642 1 T44253 transducer protein  
34 148 6.5 729 2 E70803 hypothethical prote  
35 148 6.5 881 2 S56032 probable membrane  
36 148 6.5 1556 2 A60988 saliva-interacting  
37 147.5 6.5 1147 2 T35781 hypoethical prote  
38 147.5 6.5 1302 1 JC6009 surface-located me  
39 147 6.5 778 2 T48897 transducer protein  
40 147 6.5 5327 2 T13564 microtubule-associ  
41 146.5 6.5 446 2 H83098 hypothethical prote  
42 146.5 6.5 571 2 D86164 hypothethical prote  
43 146.5 6.5 643 2 H84305 Htr8 transducer [i  
44 146 6.5 1714 1 S18644 multifunctional am  
45 145.5 6.4 628 2 F84219 Htr16 transducer [

ALIGNMENTS

RESULT 1

I40729

hypothetical 76k protein - Chlamydothila pneumoniae (strain AR39)

C;Species: Chlamydothila pneumoniae, Chlamydia pneumoniae

C;Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 31-Mar-2000

C;Accession: I40729

R;Peréz-Melgosa, M.; Kuo, C.

Infect. Immun. 62, 880-886, 1994

A;Title: Isolation and characterization of a gene encoding a Chlamydia pneumoniae

A;Reference number: I40729; MUID:94156481

A;Accession: I40729

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-715 <RES>

A;Cross-references: GB:I23921; NID:g435961; PIDN:AAA23117.1; PID:g435962

A;Experimental source: strain AR-39

C;Comment: This is the hypothetical translation of a sequence that was reported as t

Query Match 97.9%; Score 2214; DB 2; Length 715;

Best Local Similarity 98.9%; Pred. No. 1.4e-110;

Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MVNPIGPGIDETERTPPADLSAAGLEASAAKSAEAOIRIAGAEAKPKESKTSVERNSI	60
Db	257	LVNPIGPGIDETERTPPADLSAAGLEASAAKSAEAOIRIAGAEAKPKESKTSVERNSI	316
Qy	61	LRSAVNALMSLADKLGIASSNSSSTSRSDVDSTTATATPTPPPTSDDYKTQATAYDT	120
Db	317	LRSAVNALMSLADKLGIASSNSSSTSRSDVDSTTATATPTPPPTSDDYKTQATAYDT	376
Qy	121	IFTSTSLADIQAALVSLQDAVTNKTAADEETAIAAEWETKNADATKVGQAITEAKY	180
Db	377	IFTSTSLADIQAALVSLQDAVTNKTAADEETAIAAEWETKNADATKVGQAITEAKY	436
Qy	181	ASDNOAQLDLSLGLKTSFLLQTLQSVANNKAAELKEMQDNVPVPGKTPALQAQSLVD	240
Db	437	ASDNOAQLDLSLGLKTSFLLQTLQSVANNKAAELKEMQDNVPVPGKTPALQAQSLVD	496
Qy	241	QTDATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAKAAIATAKTOIAEAQ	300
Db	497	QTDATATQIEKDGNAIRDAYFAGONASGAVENAKSNNSISNIDSAKAAIATAKTOIAEAQ	556
Qy	301	KKFPDPSPLQEAQOMVQAEDKLNKIPADGSDVPNPGCTTVGGSKQOGSSIGSRVSMLL	360
Db	557	KKFPDPSPLQEAQOMVQAEDKLNKIPADGSDVPNPGCTTVGGSKQOGSSIGSRVSMLL	616
Qy	361	DDAENETASILMSGFRQMIHMFNTENPDQSAQAQELAAQARAAGDSDSAAALADAQK	420
Db	617	DDAENETASILMSGFRQMIHMFNTENPDQSAQAQELAAQARAAGDSDSAAALADAQK	676
Qy	421	ALEALKAGKQOQOGLNLGQIASAAVVSAGVLPLOQVL	459



Db 677 ALEAALGKAGQGGILNALGQIASAAVVSAGVLPQQVL 715  
|||||  
RESULT 2  
D72042 conserved hypothetical protein CP0018 [imported] - Chlamydia pneumoniae (strains CWI  
N;Alternate names: chlpn 76 kda homolog\_1 (ct622); hypothetical protein CP0728  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: D72042; D81623  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: D72042  
A:Molecule type: DNA  
A:Residues: 1-651 <ARN>  
A:Cross-references: GB:AE001654; GB:AE001363; NID:g4377031; PID:RAD18867.1; PID:g437703  
A:Experimental source: strain CWI029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: D81623  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-651 <REA>  
A:Cross-references: GB:AE002161; NID:g7188948; PID:AAF37914.1; PID:g718894  
A:Experimental source: strain AR39, HL cells  
C:Comment: This sequence was originally identified as homologous to part of a sequence  
PIR:H71490).  
C:Genetics:  
A:Gene: CPn0728; CP0018  
Query Match 96.8%; Score 2188; DB 2; Length 651;  
Best Local Similarity 98.9%; Pred. No. 3e-109;  
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MWNPIGPGPIDETERTPPADLSAQGLEASAANKSAEQAQIAGAEAKPKESKTSVERWSI 60  
Db 1 MWNPIGPGPIDETERTPPADLSAQGLEASAANKSAEQAQIAGAEAKPKESKTSVERWSI 60  
QY 61 LRSVAVNLSADKLGIASSNSSTSRSDVSTTATPTPPPTSDDYKTAQATAYDT 120  
Db 61 LRSVAVNLSADKLGIASSNSSTSRSDVSTTATPTPPPTSDDYKTAQATAYDT 120  
QY 121 IFTSTSLADIQAALVSLQDAVTNKTATDETAIAAEWETKNADAIVKGAQITELAKY 180  
Db 121 IFTSTSLADIQAALVSLQDAVTNKTATDETAIAAEWETKNADAIVKGAQITELAKY 180  
QY 181 ASDNQAILDSLGKLTFFDLLOALLOSVAANNKAAELLEMQDNPVPGKTPAIAQSLVD 240  
Db 181 ASDNQAILDSLGKLTFFDLLOALLOSVAANNKAAELLEMQDNPVPGKTPAIAQSLVD 240  
QY 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAAIATAKTQIAEAQ 300  
Db 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAAIATAKTQIAEAQ 300  
QY 301 KKFPDPSPILOEAQWVIOAEKDLKNIKPADGSDVPNPGTTVGSKQOGSSIGSIRVSMIL 360  
Db 301 KKFPDPSPILOEAQWVIOAEKDLKNIKPADGSDVPNPGTTVGSKQOGSSIGSIRVSMIL 360  
QY 361 DDENETASTILMSGFQMIHMFNTENPDQAQOELAAQAAKAAAGDSSAAALADAQK 420  
Db 361 DDENETASTILMSGFQMIHMFNTENPDQAQOELAAQAAKAAAGDSSAAALADAQK 420  
QY 421 ALEAALGKAGQGGILNALGQIASAAVVSAGVLP 454  
Db 421 ALEAALGKAGQGGILNALGQIASAAVVSAGVLP 454  
RESULT 4  
G71490 hypothetical protein CT622 - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
N:Alternate names: chlpn 76kda homolog CT622  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: G71490  
R:Stevens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitc  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia  
A:Reference number: A71570; MUID:95000809  
A:Accession: G71490  
A:Molecule type: DNA  
A:Residues: 1-647 <ARN>  
A:Cross-references: GB:AE001333; NID:g3329068; PIDN:AAAC68226.1; PID:g3

Db 677 ALEAALGKAGQGGILNALGQIASAAVVSAGVLPQQVL 715  
|||||  
RESULT 2  
D72042 conserved hypothetical protein CP0018 [imported] - Chlamydia pneumoniae (strains CWI  
N;Alternate names: chlpn 76 kda homolog\_1 (ct622); hypothetical protein CP0728  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: D72042; D81623  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: D72042  
A:Molecule type: DNA  
A:Residues: 1-651 <ARN>  
A:Cross-references: GB:AE001654; GB:AE001363; NID:g4377031; PID:RAD18867.1; PID:g437703  
A:Experimental source: strain CWI029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: D81623  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-651 <REA>  
A:Cross-references: GB:AE002161; NID:g7188948; PID:AAF37914.1; PID:g718894  
A:Experimental source: strain AR39, HL cells  
C:Comment: This sequence was originally identified as homologous to part of a sequence  
PIR:H71490).  
C:Genetics:  
A:Gene: CPn0728; CP0018  
Query Match 96.8%; Score 2188; DB 2; Length 651;  
Best Local Similarity 98.9%; Pred. No. 3e-109;  
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MWNPIGPGPIDETERTPPADLSAQGLEASAANKSAEQAQIAGAEAKPKESKTSVERWSI 60  
Db 1 MWNPIGPGPIDETERTPPADLSAQGLEASAANKSAEQAQIAGAEAKPKESKTSVERWSI 60  
QY 61 LRSVAVNLSADKLGIASSNSSTSRSDVSTTATPTPPPTSDDYKTAQATAYDT 120  
Db 61 LRSVAVNLSADKLGIASSNSSTSRSDVSTTATPTPPPTSDDYKTAQATAYDT 120  
QY 121 IFTSTSLADIQAALVSLQDAVTNKTATDETAIAAEWETKNADAIVKGAQITELAKY 180  
Db 121 IFTSTSLADIQAALVSLQDAVTNKTATDETAIAAEWETKNADAIVKGAQITELAKY 180  
QY 181 ASDNQAILDSLGKLTFFDLLOALLOSVAANNKAAELLEMQDNPVPGKTPAIAQSLVD 240  
Db 181 ASDNQAILDSLGKLTFFDLLOALLOSVAANNKAAELLEMQDNPVPGKTPAIAQSLVD 240  
QY 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAAIATAKTQIAEAQ 300  
Db 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAAIATAKTQIAEAQ 300  
QY 301 KKFPDPSPILOEAQWVIOAEKDLKNIKPADGSDVPNPGTTVGSKQOGSSIGSIRVSMIL 360  
Db 301 KKFPDPSPILOEAQWVIOAEKDLKNIKPADGSDVPNPGTTVGSKQOGSSIGSIRVSMIL 360  
QY 361 DDENETASTILMSGFQMIHMFNTENPDQAQOELAAQAAKAAAGDSSAAALADAQK 420  
Db 361 DDENETASTILMSGFQMIHMFNTENPDQAQOELAAQAAKAAAGDSSAAALADAQK 420  
QY 421 ALEAALGKAGQGGILNALGQIASAAVVSAGVLP 454  
Db 421 ALEAALGKAGQGGILNALGQIASAAVVSAGVLP 454



A:Experimental source: serotype D, strain UW-3/Cx  
C:Comment: This sequence was originally identified as homologous to part of a sequence  
PIR:E7042.  
C:Genetics:  
A:Gene: CT622

Query Match 20.0%; Score 451.5; DB 2: Length 647;  
Best Local Similarity 29.6%; Pred. No. 6.7e-17;  
Matches 138; Conservative 85; Mismatches 192; Indels 51; Gaps 14;  
QY 2 VNPICPIDEVTPPADLSAQGLEASAAKSAEQAORTAGAEAKPKESKTSVERWSIL 61  
DB 15 MPIINGQI-----ASNSETKESTKESEA-----SPASSSVSWSL 52  
QY 62 RSVAVALMSLADKLGIASSNSSTSRSDVDTTATPTPPPTDDY---KTOAQTV 118  
DB 53 SAKHALISLRD--ALLNKNSSPTDSL-OLEASTSTSTVTRVARDYNEAKSNFDTAK 108  
QY 119 DTFTSTSLADIAQALVSLQDVAITNKDTAATDEETAIAAEWETKNADAIVGAQITELA 178  
DB 109 SGLNATTLLAEVETKMAQLMADLMERLAKQAEVTRIKALQEKQEVID---KLNQLV 165  
QY 179 KYASDNOAILSLGKLTSLDLOQTALLQSVANNKKAELLKEMQDNPVPGTKPTAAQSL 238  
DB 166 KLEQONQTLKETLTITTTDSADQIPAINSOLEINKNSADQIKDLEGONI---SYEAVLTNA 222  
QY 239 VQOTDATATQIEKDNAGIDAYFAGNAGSVAENAKSNNSINIDSKAAIATAKTOIAE 298  
DB 223 GEVIRASSPAGIKLQALQISVDAGQSOAAVLQAOQNSPDNIAATKLIIDAAETKVN 282  
QY 299 AOKK---FPDPSILOAEQMVQAEKDLKNIPADGSDVPNPGTTVGGSKQOQSSIGSI- 354  
DB 283 LKQHTGLTDLSPVKKAEQISOAKDIQEIKPS-GSDIPIVGPS-GSAASAGSAGVGLK 340  
QY 355 -----RVSMILLDDAENETASILMSGFROMIHMFNTENPDQSOAQLEAAQARA- 405  
DB 341 SSNNSGRIILLDDVDNEMAAATAMQFRSMIEQFNVPNNPATAKELOAMEAQLTMSDQLV 400  
QY 406 AGDSDAAAALADAQKALEAKGAGQOQOQILNALGOITASAAYVSAG 451  
DB 401 GADGELPAEIQAKDALAQAL-KQPTDGLATAMGOVAFAAKVG 445  
RESULT 5  
C85693  
probable membrane protein of prophage CP-933X Z1918 [imported] - Escherichia coli (strain  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: C85693  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C85693  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-973 <STO>  
A:Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UWCP:Z19  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z1918

Query Match 8.9%; Score 201; DB 2: Length 973;  
Best Local Similarity 22.9%; Pred. No. 0.0024;  
Matches 120; Conservative 64; Mismatches 210; Indels 130; Gaps 14;  
QY 26 LEASANKSAEQAORTAGAEAKPKESKTSVERWSILRSVAVALMSLADKLGIASSNSSSS 85  
DB 112 VEEVARNASAVAQNTAAAKKSADASTSAREATHATDAADSARAASTSAGQAAS-SAOS 170

QY 86 TSSADVDTTATPTPPPTSDDYKTQAOATYDTTSTLADIAQALVSLQDVAITNK 145  
DB 171 ASSACTASTKATEASKSAAAESSKSAATSAAGAKTSET-----NAAVSQOASATSAS 225  
QY 146 DTAATDEETAIAAEWETKNADAIVGAQITELAKYASDNOAILDLSLGLTKTSFLLQALL 205  
DB 226 TATTKASEAASSA---RDASASKEAKSSETSAASSASAAASSAATAAGNSAKAAKTS-- 279  
QY 206 QSVANNKKAELLKEMQDNPVPGTKPTA-----IAQSLVDQTDATATQIEK- 252  
DB 280 ---ETNAKSETAEOSASAAAGSKTAALASASAASSTAGQASASATAAGKSAESAASSA 336  
QY 253 -----GNAIGDAYFAGNAGSA---VENAKSNNSIS----- 280  
DB 337 STATTKAGEATEQASAAASASSAAKTSSETNAKASETSAESKTTAAASSASSASSASSAS 396  
QY 281 -----NIDSKAAIATAKTOIAE-----AOKKFPDPSILOAEQMVQAEKDLK 324  
DB 397 ASKDEATROASAKSSATTASTKATEAGSATAAAGSKSTAESAATRAETAARAEADIAS 456  
QY 325 NIKPADGSDVPNPGTTVGGSKQOQSSIGSTRVSMILLDDAENETASILMSG----- 374  
DB 457 AVALEDAS-----TTKGIVQLSSATNSTSESAAATPKAVKAAAYELANGKYTAQDATTA 510  
QY 375 FROMIHMFNTENPDQSOAQLEAAQARAAGD-----D 409  
DB 511 QKGIVQLSSATNSTSEM---LAATPKSVKAAAYDLANGKYTAQDATTAQKGIQLSSATN 566  
QY 410 SAAALADAKALEALGAGQOQOQILNALGOITASAAYVSAGVL 453  
DB 567 SASSETLAATPKAVKAAAND-----NANGRVPSARKVNGKAL 601  
RESULT 6  
G64887  
probable tail fiber protein gp37 - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 21-Jul-2000  
C:Accession: G64887; T09189  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A:Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: G64887  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1122 <BLAT>  
A:Cross-references: GB:AE0000234; GB:U000096; NID:g1787633; PIDN:AAC74454.1; PID:g17876  
A:Experimental source: strain K-12, substrain MG1655  
R:Albal, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai,  
M.; Motomura, K.; Nakade, S.; Nakamura, Y.; Nashimoto, H.; Nishio, Y.; Oshima, T.; Sai  
moto, Y.; Horiuchi, T.  
DNA Res. 3, 363-377, 1996  
A:Title: A 570-Kb DNA sequence of the Escherichia coli K-12 genome corresponding to t  
A:Reference number: Z16603; MUID:97251357  
A:Accession: T09189  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 3-1122 <ATB>  
A:Cross-references: GB:AE0000234; GB:U000096; NID:g1787633; PIDN:AAC74454.1; PID:g17876  
Query Match 8.3%; Score 188; DB 2: Length 1122;  
Best Local Similarity 23.0%; Pred. No. 0.014;  
Matches 103; Conservative 68; Mismatches 203; Indels 74; Gaps 13;  
QY 8 GPIDEFETPPA---DLSAQGLEASAAKSAEQAORTAGAEAKPKESKTSVERWSILRS 63  
DB 93 GAMTEDDARPEALRRFELM---VEEVARNASAVAQNTAAAKKSADASTSAREATHAAD 149  
QY 64 AVNALMSLADKLGIASSNSSTSRSDVDTTATPTPPPTSDDYKTQAOATYDTT 123



Db 150 AADSAARAASSTAGQAAS--SAQSASSAGTASTKATEASKSAAABESSKAAATSAKAAT 208  
QY 124 STSLADIOALVSLQDAVNIKDTATDEETALAEWETKNADAIVKGAQITELAKYASD 183  
Db 209 SETNAS-----ASLQSAATASATTKASEAATSA-----RDAASKEAAKSETNASSA 259  
QY 184 NOAILDSLGLTSLFDLQTLQALQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVDQTD 243  
Db 260 SSAASATAAGNSAKAATIS-----ETNARSETAAGQASASAAAGSKTAAASASASATIS 314  
QY 244 ATATQTEKQGNAGIDGAYFAGQNASGAVENAKSNNSISNIDSAAKAIATAKTQIAEAKKF 303  
Db 315 A-----GQASASATAAGKSA-----ESAASSASTATTKAGEATEO- 349  
QY 304 PDSPILOEAFOMVIOAEKDLKNIKPADGSDVPNPGTTVGSKSQQSGSIGSRVSMLLDDA 363  
Db 350 -----ASAARSASAKTSETNKASETSAESKTAASASASASASASASK--DEA 402  
QY 364 ENETASILSGFQMIHMFNTENPDSSQAQQLAAQA-----RAAKAAGDSDSAAALAD 417  
Db 403 TRQASAKSSATTASTKATEAGSATAAQAQSKSTAESAATRAETAAKRAEDIASAVALD 462  
QY 418 AQKALEALGKACQOQGIINALGQIASA 445  
Db 463 ASTT-----KRGIV-----QLSSA 476

RESULT 7  
T31110  
extracellular matrix binding protein - Abiotrophia defectiva (fragment)  
C:Species: Abiotrophia defectiva  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T31110  
R:Manganelli, R.; van de Rij, I.  
Infect. Immun. 67, 50-56, 1999  
A:Title: Cloning and characterization of emb, a gene encoding the major adhesin of Strep.  
A:Reference number: Z20988; MUID:99081722  
A:Accession: T31110  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2055 <MAN>  
A:Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AAD03320.1  
C:Genetics:  
A:Gene: emb

Query Match 8.1%; Score 182.5; DB 2; Length 2055;  
Best Local Similarity 22.9%; Pred. No. 0.06;  
Matches 109; Conservative 76; Mismatches 189; Indels 101; Gaps 21;  
QY 19 ADLSAQGLEASAAKSAEQAQRIAGAEA-----KPKESKTSVERWSILRSVAVNALMSLD 73  
Db 1371 ATAKNAIDAATSNDDKATQNECTQAINAVPQTTPRAKTD-----AKNAVQQAAD 1420  
QY 74 KLGIASSNSSST-----SRSDVDSTTATAPTPPTSDDYKTQATAYDTITSTSLA 128  
Db 1421 RKXDAIENDPNLTREEKVAAKAVDA-----EAKKAKDAIDAATNSA 1462  
QY 129 DIQAALVSLQDAVNIKDTATDEETALAEWETKNADAIVKGAQITELAKYASDNOA 186  
Db 1463 DVTAQKNEGKTAINDVPQTPTAKTDKNAVTOQAADAKK--DAIEKDPNLTREEKDA--KA 1519  
QY 187 ILDSLGLTSLFDLQTLQALQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVDQTD 244  
Db 1520 KVDAEAKK--DAIDAATSNADVTAKQNEGTKAINDVPQPT--AKTDA--KNAVTOQAADA 1575  
QY 245 TATQIEKQGNAGIDGAYF-----GQNASGAVENAKSN-----NSISNI 282  
Db 1576 KKAIEKDPNLTREEKDAAKAVDAEAKKAKDAIDAATSNADVTAKQDAGKNAINAVPQT 1635  
QY 283 DSAKAAATATAKTQIAEAKK--KFPDSPILOEAEQMVIOAEKDLKNIKPADGSDVPNPGTTV 341  
Db 1636 PTAKTDKNAVTOQAADKDAIENDANLTREEKDAAKAVDAEATKAKNAIDAATSNADV 1695

QY 342 GGSKOQGS-SIGSI-RVSMLLDDAEN---ETASILSGFQMIHMFNTEN-----PDSQ 390  
Db 1696 TAKQNEGKTAINDVPQTPTAKTDKNAVDOATDKKSAI-----ENDPALTREEKD 1746  
QY 391 AAQQLAAQARAARAAAGDDSAALADAAQKALEAALGKAGQOQGIINALGQIASA 445  
Db 1747 AAKAKVDAEATKAKNAIDAATSNADVTAKQ-----DAG--KNAINAVPQTPTA 1792  
RESULT 8  
S33441  
EF protein - Streptococcus suis  
C:Species: Streptococcus suis  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 15-Oct-1999  
C:Accession: S33441  
R:Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.  
submitted to the EMBL Data Library, May 1993  
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are abs  
A:Reference number: S33441  
A:Accession: S33441  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1822 <SMI>  
A:Cross-references: EMBL:X71880; NID:g298031; PIDN:CAA50714.1; PID:g298032

Query Match 7.9%; Score 177.5; DB 2; Length 1822;  
Best Local Similarity 23.6%; Pred. No. 0.095;  
Matches 116; Conservative 77; Mismatches 195; Indels 103; Gaps 21;  
QY 13 TERTPADLSAQGLEASAAKSAEQAQRIAGAEAKPRES-KTDSVERWSIL-RSAVNALMS 70  
Db 1201 TEKAKAELAGEKSLDTGKEARDAVELAKDELAKELRTBEEETKIVEKLAEDTRKA 1260  
QY 71 LADKLGIASSNSSSTSRGAD--VDSTTATAPTPPTSDDYKTQATAYDTITSTSLAD 129  
Db 1261 IEDNPNLSDEDKQAEIKKLTDAVAKTLATI-----RDNADKRTQEAKEA-----QALAD 1309  
QY 130 IQAALVSLQDAVNIKDTATDEETALAEWETKNADAIVKGAQITELAKYASDNOAI 187  
Db 1310 LEKA-----KETQKIADKKAIDRLTILVKDGELEATKQDAKNAKIAKAAKAIASNP 1364  
QY 188 LDSLGLTSLFDLQTLQALQSVANNKA-----AELLKEMQDNVPVPGKTPAIAQSLVD 240  
Db 1365 LTDAEKKTFTD-----AVDAEAVKANDAISAATSPADVQKE--EDAGVA-----AIAEDVLD 1414  
QY 241 OTDATA-TQIEKQGNAGIDGAYFAGQNASG-----VENAKSNNSISNIDSAKAAI 289  
Db 1415 AAKODAKNKIAKDAKAAKAAKEAIGSNPLTDAEKKTTFTDAVDAEAVAKANDAIS----- 1465  
QY 290 ATAKTQIAEAKKFPDSPILOEAEQMVIOAEKDLKNIKPADGSDV-----PNPGTT--- 340  
Db 1466 --AATSPADVQKE--EDAGVAAIAEDVLDAAKQDAKN-KIAKESDAKSAIDANPLTDAE 1521  
QY 341 -----VGGSKQOQSGSIGSRVSM-----LDDAENETASILM 372  
Db 1522 KESAKKAVDAKAAKAAATDAIDASTSPVEAQSAEDKGVGSIADQVLDAAKQAKNKIAKEVA 1581  
QY 373 SGFQMIHMFNTENPDSSQAQQLAAQARAARAAAGDDSAALADAAQKALEAALGKAGQ 432  
Db 1582 AAKEAIDANPLNLSDAEAKSKAVDAKATTAID--ASTSPVEAQSAEDKGVGSI--R 1637  
QY 433 QGILNALGQIA 443  
Db 1638 QDVLDAAKQDA 1648

RESULT 9  
E84327  
Htr7 transducer [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001



C:Accession: E84327  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, R.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon  
Jung, K.H.; Alam, M.; Freitas T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: E84327  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-545 <STO>  
A:Cross-references: GB:AE004437; NID:g10581214; PIDN:AAG19985.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: htrI  
C:Superfamily: Halobacterium salinarum transducer protein htrI

Query Match 7.7%; Score 174.5; DB 2; Length 545;  
Best Local Similarity 21.7%; Pred. No. 0.029;  
Matches 100; Conservative 79; Mismatches 192; Indels 89; Gaps 17;

Qy 12 ETERPPADLSAOGLEASAAKSAQAORIAAGAEAKPKESKTDVSVERWILRSVAVNALMSL 71  
Db 118 EEERA-----EAERAREAKQQAER---QTAEASAKQDARERSAEIQLAADLESQ 169  
Qy 72 ADKLGIASSNSSSTSRSDVSTTATPTPPPT-SDYKTQAQAYDTI-----FTST 125  
Db 170 ATEVG-ATLEAASDGLTARVDATTDNAEIAEVATVNDMLTTMERTIDEIOGFSTNVTT 228  
Qy 126 SLADIQAALVSLQDAVTNIKOT-----NATDEE-----TAIAAEWETKNADAIVKGAQITE 176  
Db 229 ASRETAGAKETQDASQTVSEVQEIAGTDDOREQLSEAEEMDSYATVEEVAATAQS 288  
Qy 177 LAKYASDNOAILDSLGKLT-----SFDLLQTLALQSVAANNKAAELKEMODNPVPGK 230  
Db 289 VADTAADTTDVA-TAGKQTAEDAIDAIVQETMOTTVANVDALEDLLEIDT----- 340  
Qy 231 TPATQAQSLVQDTAT-----ATQIEKDGNAIGD--AYFAGQNASGAVENAKSNNS 278  
Db 341 ---IAELISDAEQTNMLALNANIEAARAGSGGSGDGFVAVDEVKELATESQSAKD 397  
Qy 279 ISN-IDSAKAATAKTQIAEAKKFPDPSPILOEAEQMVIAEKDLKNIKPADGSDVNP 337  
Db 398 IAELEEVSQATTVVEI-----RVAEQRVNCGAAVEETVDAFGAVTENI 444  
Qy 338 GTTVGGSKQGGSSIGSIRVSMLLDDAENETASILMSGFRQIMHMFNTENPDPSQAAQBELA 397  
Db 445 QETTDGVQE-----ISQAMDEQAQSRVSVSS-----VDDIATISQATA 483  
Qy 398 AQARAKAAGDSDRAALADAKALEAALGKAGQOQGILN 437  
Db 484 DRAENVSAASEEQ-TASITEVTSLSQSLAAQAQDTLEDRLN 522

RESULT 10  
T46811  
halobacterial transducer protein V [imported] - Halobacterium salinarum  
C:Species: Halobacterium salinarum  
C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 15-Sep-2000  
C:Accession: T46811  
R:Rudolph, J.; Nordmann, B.; Storch, K.F.; Gruenberg, H.; Rodewald, K.; Oosterheilt, D.  
FEMS Microbiol. Lett. 139, 161-168, 1996  
A:Title: A family of halobacterial transducer proteins.  
A:Reference number: 224094; MUID:96275896  
A:Accession: T46811  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-545 <RUD>  
A:Cross-references: EMBL:X95589; NID:g1435130; PIDN:CAA64842.1; PID:g1435132  
C:Genetics:  
A:Gene: htp  
C:Superfamily: Halobacterium salinarum transducer protein htrI

Query Match 7.7%; Score 174.5; DB 2; Length 545;  
Best Local Similarity 21.7%; Pred. No. 0.029;  
Matches 100; Conservative 79; Mismatches 192; Indels 89; Gaps 17;

Qy 12 ETERPPADLSAOGLEASAAKSAQAORIAAGAEAKPKESKTDVSVERWILRSVAVNALMSL 71  
Db 118 EEERA-----EAERAREAKQQAER---QTAEASAKQDARERSAEIQLAADLESQ 169  
Qy 72 ADKLGIASSNSSSTSRSDVSTTATPTPPPT-SDYKTQAQAYDTI-----FTST 125  
Db 170 ATEVG-ATLEAASDGLTARVDATTDNAEIAEVATVNDMLTTMERTIDEIOGFSTNVTT 228  
Qy 126 SLADIQAALVSLQDAVTNIKOT-----NATDEE-----TAIAAEWETKNADAIVKGAQITE 176  
Db 229 ASRETAGAKETQDASQTVSEVQEIAGTDDOREQLSEAEEMDSYATVEEVAATAQS 288  
Qy 177 LAKYASDNOAILDSLGKLT-----SFDLLQTLALQSVAANNKAAELKEMODNPVPGK 230  
Db 289 VADTAADTTDVA-TAGKQTAEDAIDAIVQETMOTTVANVDALEDLLEIDT----- 340  
Qy 231 TPATQAQSLVQDTAT-----ATQIEKDGNAIGD--AYFAGQNASGAVENAKSNNS 278  
Db 341 ---IAELISDAEQTNMLALNANIEAARAGSGGSGDGFVAVDEVKELATESQSAKD 397  
Qy 279 ISN-IDSAKAATAKTQIAEAKKFPDPSPILOEAEQMVIAEKDLKNIKPADGSDVNP 337  
Db 398 IAELEEVSQATTVVEI-----RVAEQRVNCGAAVEETVDAFGAVTENI 444  
Qy 338 GTTVGGSKQGGSSIGSIRVSMLLDDAENETASILMSGFRQIMHMFNTENPDPSQAAQBELA 397  
Db 445 QETTDGVQE-----ISQAMDEQAQSRVSVSS-----VDDIATISQATA 483  
Qy 398 AQARAKAAGDSDRAALADAKALEAALGKAGQOQGILN 437  
Db 484 DRAENVSAASEEQ-TASITEVTSLSQSLAAQAQDTLEDRLN 522

RESULT 11  
F84194  
HtrI4 transducer [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: F84194  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, R.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jung, K.H.; Alam, M.; Freitas T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: F84194  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-627 <STO>  
A:Cross-references: GB:AE004437; NID:g10579976; PIDN:AAG18922.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: htrI4  
C:Superfamily: Halobacterium salinarum transducer protein htrI

Query Match 7.6%; Score 171; DB 2; Length 627;  
Best Local Similarity 23.4%; Pred. No. 0.053;  
Matches 120; Conservative 87; Mismatches 184; Indels 122; Gaps 27;

Qy 17 PPADLSAOGLEASAAKSAQAORIAAGAEAKPK-----ESKTDVSVERWILRSVAVNALMSL 71  
Db 145 PALDESVPAGGESITEMADSLEAYTAEDKTALEHQAELEKRSOLRALVDALSEA 204  
Qy 72 ADKLGIASSNSSSTSRSDVSTTATPTPPPTSDYKTQA-----QTAYDTIFTS 124  
Db 205 TD-----AARAGDL---TATVDAAALDVTDDHRAADEVFNQLETATDISDI 249



QY 125 TSLADIQAAL-----VSLQDAVTNKTDA--ATDEETAIAAEWETKNADA 167  
Db 250 QSFDAVLAVSRITTDERVDADVADRSAAVSESVTEIADGANOQTNOLNIAAEMDTVSATV 309  
QY 168 IKVGAQITELAKYASDNOAIL-----DSLGLK-TSFDLLQALLOSVAANNKAAELLKEMQ 222  
Db 310 EEIAASANDVAKTA---QAADRDGGRGEVEETIEALRALRGESQVAETVESLAAEVE 366  
QY 223 DNPVPGKTPAIAQSLVDOTDATA--TQIE-----KGNALGDAYFAGONASGAVENAKS 275  
Db 367 R---IDGIT-ALIEDIAEETNNLALNASIEAARTGSDG--GFVAVADEVKDLAEETREQ 420  
QY 276 NNSISNIDSA-----KAAIATATQIAEAKQKPPDSP-ILQEAQMVVIOAEKDLKNKP 328  
Db 421 AADISEIVDAYTEKADASIAIGVD-AEVERKITKAGVLRDFEAV-----DEVANVNH 475  
QY 329 A--DGSVDVNPNG-----TTVGSKQOOGSSI-----GSRVSMILLDD 362  
Db 476 AVQEISDATDQAGQSVTDVVGWVEEVAVSSEETAESDTVADNAAEQTDATDEADVADQME 535  
QY 363 AENETAST--LMSCFROMIHMFNTENPDQAAQELAAQARA--AAGDDSAALADA 418  
Db 536 LAETUAAALAGMLDDF-----TVPADAGTADQOSVADDSPTAQPPAADDEPAAAVVDQP 587  
QY 419 QKALEAALGRAGAQOQOQILNALGQIASAAVVSAG 451  
Db 588 QPASDAE-----DEEGVPDSGGE--SVAVSDGG 613

RESULT 12  
T30296  
R27-2 protein - Trypanosoma cruzi  
C:Species: Trypanosoma cruzi  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 17-Mar-2000  
C:Accession: T30296  
R:Otsu, K.; Donelson, J.E.; Kirchhoff, L.V.  
Mol. Biochem. Parasitol. 57, 317-330, 1993  
A:Title: Interruption of a Trypanosoma cruzi gene encoding a protein containing 14-amino acid repeats  
A:Reference number: 220813; MUID:93165082  
A:Accession: T30296  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1128  
A:Cross-references: EMBL:L04603; NID:G385171; PID:G1256742; PIDN:AAA96494.1  
C:Superfamily: neurofilament triplet H protein

Query Match 7.4%; Score 166.5; DB:2; Length 1128;  
Best Local Similarity 23.9%; Pred. No. 0.2;  
Matches 110; Conservative 71; Mismatches 202; Indels 77; Gaps 19;

QY 10 IDERTPPADLS---AOGLEASAANKSAEAOQRIAGAE--KPKESKTDVSVERWSILRSA 64  
Db 642 VAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAE 701  
QY 65 VNALMSLADKLIGIASNSSSSSTSRSDVSTTATPTPPPTSDDYKTQATYDTIFTS 124  
Db 702 EKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEAT---KVAEAEKQKAAE-----TK 753  
QY 125 TSLADIQAALVSLQDAVTNKTDAATDEETAIAAEWETKNADA 177  
Db 754 VAEAEKQKAAEATKVAEAE--KQKAA--EATKVAEAEKQKAAEATKVAEAEKQKAAEATKV 810  
QY 178 AKYASDNOAILDSLGLKTSFDLLQALLOSVA--ANNKAAELLK-----EMQDNPVVPCKT 231  
Db 811 AEAEKQKAA---EATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVA 867  
QY 232 PAIAQSLVDQTDATATQIEKDGNAIGDAYFAGONASGAVENAKSN-----NSISNIDS 284  
Db 868 EAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEK 927  
QY 285 AKAAIATAKQIAEAKQ-KFPDPSILOEABQM-----VIOAEKDLKNKIPADGSDVPN 336

Db 928 QKAAEA---TKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQ-----KAAEATKVAE 980  
QY 337 PCTTVGSKQOOGSGSIRVSMILLDDAENETASILMSGFROMIHMFNTENPDQAAQOEL 396  
Db 981 -----EAKQKAAEATKVA-----EAKQKAAEATK-----VAEAEKQKAAEATK 1019  
QY 397 AAQARAAKAGDDGDDSAALADAOKALEAALGKAGQOQOIL 436  
Db 1020 VAEAEKQKAA--EATKVAEAEKQKAAEAKAMESQKORFL 1057

RESULT 13  
T34434  
hypothetical protein K06A9.1a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T34434  
R:Geisel, C.; Gattung, S.  
submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid K06A9.  
A:Reference number: 221525  
A:Accession: T34434  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2232  
A:Cross-references: EMBL:U80846; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a  
A:Experimental source: strain Bristol N2; clone K06A9  
C:Genetics:  
A:Gene: CESP:K06A9.1a  
A:Map position: X  
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 21/2;

Query Match 7.4%; Score 166.5; DB:2; Length 2232;  
Best Local Similarity 19.7%; Pred. No. 0.47;  
Matches 79; Conservative 70; Mismatches 174; Indels 79; Gaps 11;

QY 4 PIGPGPIDETERTPPADLSAQ---GLEASAANKSAEAOQRIAGAE--KPKESKTDVSVERWS 59  
Db 598 PSSOSPAPNTGSTTPTSTQSPSPSPMNPSSSTPTGSSQSTTITPEGSTASSPTGSTGTF 657  
QY 60 ILRSVAVNALMSL--ADKLIGIASNSSSSSTSRSDVSTTATPTPPPTSDDYKTQATA 117  
Db 658 SVATEVTSQSTVPSVPSGLTQSTNSPSPSLSPSTSCMSLTSEPPSS---TQSSGA 713  
QY 118 YDTIFT--STSLADIQAALVSLQDAVTNKTDAATDEETAIAAEWETKNADA 176  
Db 714 QSTLTTPSPNPSPQSTSLSSSTSGATTSAGT-----MTSPSQSSSVG----- 759  
QY 177 LAKYASDNOAILDSLGLKTSFDLLQALLOSVA--ANNKAAELLKEMQDNPVVPCKTPTAIAQ 236  
Db 760 -SSOGSTSPRASTTSGEMTSQGSTQT-----PGSSVSTSA 793  
QY 237 SLVDQTDATATQIEKDGNAIGDAYFAGONASGA---VENAKSNNSISNIDSAKAAIATAK 293  
Db 794 AILTSTQOVSSTNSPSTGSTRPSTVSGTSSGTVTVGSTEASTSGSVASSSPAPSTSQ 853  
QY 294 TQIAEAKQKPPDPSILOEABQMVIQAEKDLKNKIPADGSDVPNPG-----TTVGSKQOQ 347  
Db 854 N-----PNPSTSGSMTQSPYPSQSTSPVESSTTTPSPGPTTLTSTSPSPSQ 903  
QY 348 GSSGTGSRVSMILLDDAENETASILMSGFROMIHMFNTENPD 389  
Db 904 STTIGS-----TCGSTSPGISTTSEMTSQGSTQTG 936

RESULT 14  
T44938  
transducer protein htpv [similarity] - Halobacterium salinarum  
N:Alternate names: methyl-accepting taxis protein htl; transducer protein htl; trans  
C:Species: Halobacterium salinarum  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 02-Sep-2000



C:Accession: T44938  
R:Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.  
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996  
A:Title: Signal transduction in the archaeon Halobacterium salinarum is processed through a transmembrane protein  
A:Reference number: 222804; MUD:96209786  
A:Accession: T44938  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-544 <2HA>  
A:Cross-references: EMBL:U53365; NID:gl589744; PIDN:AAB17519.1; PID:gl621047  
A:Note: the source is designated as Halobacterium salinarum  
C:Genetics:  
A:Gene: htp5; htp5; htrVII; htrI  
C:Superfamily: Halobacterium salinarum transducer protein htrI  
C:Keywords: methylated amino acid; signal transduction; transmembrane protein  
F:246-502/Region: MCP signalling domain similarity

Query Match 7.1%; Score 160.5; DB 2; Length 544;  
Best Local Similarity 22.3%; Pred. No. 0.16;  
Matches 101; Conservative 74; Mismatches 183; Indels 95; Gaps 19;

Qy 12 ETERPPADLSAAGLEASAAKSAEAKPKESKTDSDVERWSILRSVAVNALMSL 71  
Db 118 EEERA-----EERAREAKQKQAEER-OTAEASQGTGTQRE---IEOLAADLESQ 168  
Qy 72 ADKLGIASSNSSSTSRSDVSTTATPTPPPT--SDDYKTAQT-----AYDTIFTS 124  
Db 169 ATEVG-ATLEASDGLTARVDATTDNAEIAEVATVNDMLTMTERTIDEIQGFSTNVT 227  
Qy 125 TS-----LADTQALVSLQDAVNTIKDTAATDEE---TAIAAEWETKNADAIVKGAQI 174  
Db 228 ASREATGAKETIQASQTVSESVQEI--AAGTDDQREQLSVAEEMDSYSAATVEEVAATA 285  
Qy 175 TALKYASDQALDLSGLT-----SFDLIQTALLOSVAANNKAAELKEMQNPVVP 228  
Db 286 QSVADTAADTTVA-TAGKQTAEDAIDALDAVQETMTQTVANVDALEDLTTFEIDD----- 339  
Qy 229 GKTPAIAQSLVDQDAT-----ATQIEKQGNATGD--AYFAGQNASGAVENAKSN 276  
Db 340 -----IAELISDIARQTNMLNLANIAEARAGSGGTGNGDFAVVAVDEVKELATESQRS 394  
Qy 277 NSISN-IDSAAKAIATATQIAEAKKFPDPI-----LQRAE 313  
Db 395 KDIAELIEVSQATATVEEIRVABQVRVNDGAAVEETVDAFGAVTENIQETDGVQEIS 454  
Qy 314 Q-----MVIQAEKDLNKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMLLDDA 363  
Db 455 QAMDEQAQSRVSVSDDIATISQATADRAEN---VSNASEEQATISITVETGSLQSLAA 511  
Qy 364 ENETASILMSGFR-QMIRMFNTENPDQAAQOE 395  
Db 512 QADTLEDRINEFRTEATGHTGERTDAPAGQSD 544

RESULT 15  
T34852

probable secreted protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T34852  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, February 1999  
A:Reference number: Z11559  
A:Accession: T34852  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1156 <OLI>  
A:Cross-references: EMBL:AL035478; PIDN:CAB36606.1; GSPDB:GN00070; SCOEDB:SC2G5.19  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC2G5.19

Query Match 7.0%; Score 158.5; DB 2; Length 1156;  
Best Local Similarity 21.2%; Pred. No. 0.54; Indels 65; Gaps 15;  
Matches 99; Conservative 75; Mismatches 227;  
Qy 12 ETERPPADLSAAGLEASAAKSAEAKPKESKTDSDVERWSILRSVAVNALMS- 70  
Db 367 DASKTKAARLAAGAGARNAAKARKAAQAATAQAAAAAGISAAATARDSAAAAQA 426  
Qy 71 --LADKLGIASSNSSSTSRSDVSTTATPTPPPTSDDYKTAQTAYDTIFTSTSLA 128  
Db 427 AVAAQASGAAGAAEAAVAAAAEADAQAARA---TKAANRAQSLAATAASAAAAARKAA 482  
Qy 129 DTQAA-LVSLQDAVNTIKDTAATDETAIAAE-WETKNADAIVKGAQITELAKYASDNOA 186  
Db 483 DSAHAHAKEKAADAADAADAAGEADDYANKAKAWA---ADSVAA-----AELAAKAVDDAR 535  
Qy 187 ILDSLGKITSFDLLQTLQSVANNKAAELKEMQNPVVPKTPAIAQSLVDQTDATA 246  
Db 536 AVEAAAREAAEAEKLAHDTQSLAEARENA--AAEAEADREARN-----AATEADRLDAQT 588  
Qy 247 TQIEKDGNAGIDAYFAGQNASCAVENAKSNNSISNIDSAKAIATATKQIAEAKKFPDS 306  
Db 589 ---KDFISRAEAATASGDTASALANGR-----KAANLISITIGTWSRAAAEY 633  
Qy 307 PIQPAEQMVIQAEKDLNKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMLLDDAENE 366  
Db 634 ALAGEDVLTWVSTDRQIAQRQDDSE---TALAANVSTEAIVANAADAIAENSDDPQS 688  
Qy 367 TASILMSGFRQ-----HMFNTENPDQAAQOEAAQARAARAKAGDDSSAAAAALADA 418  
Db 689 VRNFLTGTGHEAAALDRVRDILRLGLDNPGR-----AVKDAQAALDDGSPSALHAF 740  
Qy 419 OKAL--EAALGRKAGQOQGLNALGQIASAAVVSAGVLPLOQVLR 462  
Db 741 FRALPDKAALDDRAVILITILNTAGPYTAAAOVA-----LEGTSWNR 782

Search completed: February 7, 2002, 21:38:09  
Job time: 7098 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:42:42 ; Search time 76.51 seconds  
(without alignments)  
221.877 Million cell updates/sec

Title: US-09-391-606-9

Perfect score: 2261

Sequence: 1 MYPNPGPIDERTPPAD.....SAAVSAGVLPQLQWLIRA 463

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	188	8.3	1120	1 STFR_ECOLI	P76072 escherichia
2	174.5	7.7	545	1 HTR5_HALN1	Q48318 halobacteri
3	158	7.0	535	1 HTR1_HALN1	P33741 halobacteri
4	158	7.0	810	1 HTR4_HALN1	Q9hp84 halobacteri
5	154.5	6.8	763	1 HTR2_HALN1	Q48317 halobacteri
6	154.5	6.8	764	1 HTR2_HALN1	Q9hp81 halobacteri
7	154.5	6.8	764	1 HTR2_HALN1	P71410 halobacteri
8	152	6.7	1561	1 SPAP_STRMU	P23504 streptococc
9	151.5	6.7	641	1 HRPX_PSESY	P41501 pseudomonas
10	150.5	6.7	1528	1 SPAA_STRDO	P21979 streptococc
11	148	6.5	505	1 FLJB_SALTY	P52616 salmonella
12	148	6.5	881	1 YJH8_YEAST	P47033 saccharomyc
13	147	6.5	778	1 HTR6_HALSA	Q48319 halobacteri
14	146.5	6.5	774	1 STF_LAMBD	P03764 bacterioph
15	146	6.5	1714	1 SYEP_DROME	P28668 drosophila
16	145	6.4	705	1 CWBA_BACSU	Q02113 bacillus su
17	145	6.4	1565	1 PAC_STRMU	P11657 streptococc
18	144	6.4	1637	1 MRSP_STAAU	P80544 staphylococ
19	143.5	6.3	1306	1 MSB2_YEAST	P32334 saccharomyc
20	142.5	6.3	778	1 HTR6_HALN1	Q9hr92 halobacteri
21	139	6.1	797	1 VGLX_HSVB	P28968 equine herp
22	138	6.1	1609	1 LMG1_HUMAN	P11047 homo sapien
23	138	6.1	2843	1 APC_HUMAN	P25054 homo sapien
24	137.5	6.1	2453	1 NCRI_MOUSE	Q60974 mus musculu
25	137.5	6.1	2492	1 TALA_DICDI	P54633 dictyosteli
26	137	6.1	500	1 FLJB_SALAE	P52615 salmonella
27	137	6.1	535	1 HTR1_HALSA	P33955 halobacteri
28	137	6.1	1607	1 LMG1_MOUSE	P02468 mus musculu
29	136	6.0	2841	1 TALI_MOUSE	Q60309 mus musculu
30	136	6.0	2845	1 APC_MOUSE	Q61315 mus musculu
31	135.5	6.0	1061	1 TRC4_ECOLI	P27189 escherichia
32	135.5	6.0	1969	1 MYSA_CAEEL	P12844 caenorhabdi
33	135	6.0	1379	1 YFF9_SCHPO	O14066 schizosacch

## ALIGNMENTS

RESULT 1

```

STFR_ECOLI
ID STFR_ECOLI STANDARD; PRT; 1120 AA.
AC P76072: P77560;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE SIDE TAIL FIBER PROTEIN HOMOLOG FROM LAMBDOID PHOPHAGE RAC.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; Pubmed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; Pubmed=9097039;
RA Alba H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Kasai H., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Makino K., Waki T., Nakamura Y., Nishio Y., Oshima T., Saito N.,
RA Nakade S., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
[3]
RP SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
[4]
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000234; AAC74454.1; ALT_INIT.
CC EMBL; D90774; BAA14966.1;
CC EMBL; D90775; BAA14975.1;
CC HSP; P04002; IWFA.
CC Ecogene; EG13370; stfr.
CC InterPro; IPR000122; Chemotaxis_transducer.
CC Hypothetical protein; Fiber protein; Repeat; Complete proteome.
CC SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;

```

```

34 134 5.9 564 1 M12_STRPY P19401 streptococc
35 134 5.9 1036 1 NIT2_NEUCR P19212 neurospora
36 134 5.9 2842 1 APC_FAT P70478 rattus norv
37 132.5 5.9 573 1 YE18_YEAST P40095 saccharomyc
38 131.5 5.8 1065 1 SED4_YEAST P25365 saccharomyc
39 131.5 5.8 1411 1 TCOF_HUMAN Q13428 homo sapien
40 130.5 5.8 3726 1 TRX_DROME P20659 drosophila
41 130 5.7 475 1 MOTD_RHIME Q52964 rhizobium m
42 130 5.7 490 1 CE05_ECOLI Q47500 escherichia
43 130 5.7 1230 1 ST20_CANAL Q22212 candida alb
44 129 5.7 1500 1 SSP5_STRGN P16952 streptococc
45 128.5 5.7 534 1 HTR2_NATPH P42259 natronomona

```



HT85_HALNI	STANDARD;	PRT;	545 AA.
IC	Q48318; Q9HP85;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	HALOBACTERIAL TRANSDUCER PROTEIN V.		
GN	HT87 OR HTPV OR VNG1759G.		
OS	Halobacterium sp. (strain NRC-1), and		
OS	Halobacterium salinarium.		
OC	Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;		
OC	Halobacterium.		
OX	NCBI_TAXID=64091, 2242;		
OX	[1]		
PC	SEQUENCE FROM N.A.		
RC	STRAIN=NRC-1;		
RC	MEDLINE=20504483; PubMed=11016950;		
RX	Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,		
RA	Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,		
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,		
RA	Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,		
RA	Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,		
RA	Tsenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.H.,		
RA	Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,		
RA	Enghardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,		
RT	"Genome sequence of Halobacterium species NRC-1.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).		
RL	[2]		
RL	SEQUENCE FROM N.A.		
RC	SPECIES=H salinarium; STRAIN=S9;		
RC	MEDLINE=96275896; PubMed=8674984;		
RX	Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,		
RA	Oesterheld D.;		
RA	"A family of halobacterial transducer proteins.";		
RT	FEMS Microbiol. Lett. 139:161-168(1996).		

RESULT	3	
HTR1_HALN1		
ID	HTR1_HALN1	STANDARD; PR7; 535 AA.
AC	P33741; Q9HPF6;	
DT	01-FEB-1994 (Rel. 28, Created)	
DT	01-JUN-1994 (Rel. 29, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	SENSORY RHODOPSIN I TRANSDUCER (HTR-I) (METHYL-ACCEPTING PHOTOTAXIS	
EN	PROTEIN I) (MPP-I).	
GN	HTR1 OR HTRI OR VNG1659G.	







This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X95589; CAA64841.1; ..  
InterPro: IPR000122; Chemotaxis\_transducer.  
InterPro: IPR000658; DUF5.  
InterPro: IPR003660; HAMP.  
Pfam: PF00672; DUF5; 1.  
Pfam: PF00015; MCPsignal; 1.  
SMART; SM00304; HAMP; 2.  
SMART; SM00283; MA; 1.  
Transducer; Transmembrane.  
KW DOMAIN 1 38  
FT TRANSMEM 39 59  
FT FT 60 323  
FT DOMAIN 60 323  
FT TRANSMEM 324 344  
FT FT 345 810  
FT DOMAIN 345 810  
FT CYTOPLASMIC (POTENTIAL).  
SEQUENCE 810 AA; 85207 MW; 37B0F6046A39D9EA CRC64;  
Query Match 7.0%; Score 158; DB 1; Length 810;  
Best Local Similarity 21.8%; Pred. No. 0.28;  
Matches 101; Conservative 75; Mismatches 169; Indels 118; Gaps 22;  
QY 27 EASAAKSA-EAQRIGAAEKPKESKTDSEVERWILSRVAVNALMSLADKLGIASSNSSSS 85  
DB 398 EAAAREQATEPAQDAEAE--RERAEDAREADAEDAKADAEAL-----AAELEAQ 444  
QY 86 TSSRADVDSVTATPTPPPTSDDYKTKQAQAYDTFTSTSLADIQAALVSLQDAVTNIK 145  
DB 445 AERYSDVWAACADGDLFRMPADDTDNEMAAIAASF-NEMLAQWEHTIIDQE---FA 499  
QY 146 DTAATDEETAJAEWETKNADKAIVGAQIT----ELAYASDNQAILDSL-GKLTSFDLL 200  
DB 500 DAVATASEA-----EVGAADAERASGVSESVQRIAGAADQRNMLDTVSGEMT--DL- 551  
QY 201 QTALLOSANNKAAELLKEMQDNVPVPGKTPAIGSLVDOTDATATOIEKDGNAIGDAY 260  
DB 552 -SAAIEEVA-----ASADSVAEHSHQTA-ETARGE----- 580  
QY 261 FAGQNASGAVENAKSNNSISNIDSAKAAIATAKTOIAEAAQKFPD--SPILQEAQMVTQ 318  
DB 581 --QTAEDAERSLSVQEA--IDATVQNVNEALDDQMAEI-SEIVDLISDIAEQTNMLALN 634  
QY 319 AE-----KDLKNIKPADGSDVPNGTIVGSGKQGGSSIGSIRVSM 358  
DB 635 ANIEAARADKSGDGFVAVVEVDKLLAEETQESAGDIERRITEV--QSQTATVAEARA 592  
QY 359 LLDDAENETASILMSGFRQMI-HMFNT-----ENPDQAAQOE----- 395  
DB 693 ESMADAGIDAVEEVDVDAFTAVSDHADETDTGVQIEISDTTDDQASSTEEAVSMTEEVADLSD 752  
QY 396 -LAAQARAAKAGDDSSAAALADAQAKALEAALGKAGQOQGIILN 437  
DB 753 STAGEAQSVSRAAEQ-AASMSEISDSVESLSGQAEQLKALLS 794  
RESULT 6  
HTR2\_HALN1  
ID HTR2\_HALN1 STANDARD; PRT; 763 AA.  
AC Q9HP81;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE SENSORY RHODOPSIN II TRANSDUCER (HTR-II) (METHYL-ACCEPTING PHOTOTAXIS  
DE PROTEIN II) (MPP-II).  
GN HTR2 OR VMGL765G.  
OS Halobacterium sp. (strain NRC-1).



Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 Halobacterium.  
 NCBI\_TaxID=64091;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shroana J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlischroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1".  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY  
 CC RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT  
 CC CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AEO05080; AAG19989.1; -  
 CC InterPro: IPR000122; Chemotaxis\_transducer.  
 CC InterPro: IPR000658; DUF5.  
 CC InterPro: IPR003660; HAMP.  
 CC Pfam: PF00672; DUF5; 1.  
 CC Pfam: PF00015; MCPsignal; 1.  
 CC SMART: SM00304; HAMP; 2.  
 CC SMART: SM00283; MA; 1.  
 CC Transducer: Photoreceptor; Transmembrane; Methylation;  
 CC Complete proteome.  
 CC INIT\_MET 0 BY SIMILARITY  
 CC DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 16 36 POTENTIAL.  
 CC DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 278 297 POTENTIAL.  
 CC DOMAIN 298 763 CYTOPLASMIC (POTENTIAL).  
 CC SEQUENCE 763 AA; 78911 MW; CF7A9FF04DFE309A CRC64;  
 SQ  
 Query Match 6.8%; Score 154.5; DB 1; Length 763;  
 Best Local Similarity 22.6%; Pred. No. 0.39;  
 Matches 105; Conservative 71; Mismatches 196; Indels 93; Gaps 18;  
 Oy 8 GPIDETERTPPADLSAQGLEASANKSAE-----AQRACAEAKPKE---SKTDSVERW 58  
 Db 336 GTLAESFRSFRSDESLSSTDAERATAREDAEDAEQQRADAEAREDAARDAQETA 395  
 Oy 59 SILRSV-----NALMSLADKLGIASSNSSSTSRSA-----DVIDSTTATAPT 101  
 Db 396 RALESAAADYEALTAVAD--GDLTRRVYDASRDHDAMARIGHALNDMLDIETSVAAATA 453  
 Oy 102 PPPPTSDYKTAQATAYDTITFTSLADIQAALVSLQDAVNTKDTAETDETAIAEWE 161  
 Db 454 FSDHVSDAARQVEADAGDAIDAGT---DVSFVAVDEISDGAET-----QTDRLHEVAGEVD 505  
 Oy 162 TKNDATAIKVGAQITELAKYASDNOAILDLSGLKLTSLFDLLOTALLOSVAANNKAPELLKEM 221  
 Db 506 DLSASAEVETVASLADTAQAASAVDD--GRQATEDRVER--MDDVADREARADAMA 562  
 Oy 222 QDNVPVP--GKTPAIAQSLVDQTDATQIEKDGNAIGDAYFAGONASGAVENA----- 273  
 Db 563 LDSEADIGEIVDVADIADQTNMLAL-----NASIEARTGADGDFVAVADEVKTALA 616

274 -KSNNSISNIDSAAKATATQTAIAEQKFPDPSPILQEAQMVIQAEKDLKNKIPADGS 332  
 Db 617 EESRDAAEEDIESRLALQGVSDVADEMRASTDT--VSDGRATVGDAAATALDDV----- 668  
 Oy 1333 DVPNPGTTVGSKQKGSGISIRVSMLLDDAENETASTILMSGFQMIHMENTENPDQAA 392  
 Db 669 -----VSFVADTDTAAGEIRAA---TDQAHAASRVASAVDEV-----AGI 706  
 Oy 393 QOELAAQARA-AKAAGD-----DSAAALADAQAQKALEAALGK 428  
 Db 707 SQETAQAQAVADSAAATQTDTLSSVDDAAADLADRAAALDDLLAE 751  
 RESULT 7  
 HTR2\_HALSA STANDARD: PRT; 764 AA.  
 ID HTR2\_HALSA  
 AC P71410;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SENSORY RHODOPSIN II TRANSDUCER (HTR-II) (METHYL-ACCEPTING PHOTOTAXIS  
 DE PROTEIN II) (MPP-II).  
 DE HTR2 OR HTRII.  
 OS Halobacterium salinarum.  
 OC Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 OC NCBI\_TaxID=2242;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FLX15;  
 RX MEDLINE=96323203; PubMed=8710852;  
 RA Zhang W., Brooun A., Mueller M.M., Alam M.;  
 RT "The primary structures of the Archaeon Halobacterium salinarum blue  
 RT light receptor sensory rhodopsin II and its transducer, a methyl-  
 RL accepting protein".  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:8230-8235(1996).  
 CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR  
 CC SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO  
 CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.  
 CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U62676; AAC44369.1; -  
 CC InterPro: IPR000122; Chemotaxis\_transducer.  
 CC InterPro: IPR000658; DUF5.  
 CC InterPro: IPR003660; HAMP.  
 CC Pfam: PF00672; DUF5; 1.  
 CC Pfam: PF00015; MCPsignal; 1.  
 CC SMART: SM00304; HAMP; 2.  
 CC SMART: SM00283; MA; 1.  
 CC Transducer: Photoreceptor; Transmembrane; Methylation;  
 CC BY SIMILARITY  
 CC INIT\_MET 0  
 CC DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 16 36 POTENTIAL.  
 CC DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 278 298 POTENTIAL.  
 CC DOMAIN 299 764 CYTOPLASMIC (POTENTIAL).  
 CC SEQUENCE 764 AA; 79187 MW; 1E0D7B4E460FC588 CRC64;  
 SQ

Query Match 6.8%; Score 154.5; DB 1; Length 764;  
 Best Local Similarity 22.6%; Pred. No. 0.39;  
 Matches 105; Conservative 71; Mismatches 196; Indels 93; Gaps 18;  
 Oy 8 GPIDETERTPPADLSAQGLEASANKSAE-----AQRACAEAKPKE---SKTDSVERW 58  
 Db 563 LDSEADIGEIVDVADIADQTNMLAL-----NASIEARTGADGDFVAVADEVKTALA 616



Db 337 GTLAESFRMRDLSLESLTDAERATARAEDAREDAEQORADAEAAAREDAEAAKDAQETA 396

QY 59 SILRSVAV-----NALMSLADKLGIASSNSSSTSRSA-----DVSSTTAPT 101

Db 397 RALESAADYEALUTAVD--GDLTRVYDASRDHDAMARIGHALNDMLDDIETSVAAATA 454

QY 102 PPPPTSDYKTAQTQATYTTFTSTSLADIAQALVSLDQAVTNKTDATDEETAEIAEWE 161

Db 455 FSDHVSAAQORVEADAGDAIDACT--DVSSTAVDEISDGAETE-----QDRLHEVAGEVD 506

QY 162 TKNADAKVGAQITELAKYASNDQAILDSGLKTSFLLQTLALLOSANNKAAELKEM 221

Db 507 DLSASAEVETASLASLATAGQASAVDD--GROATEDAVET--MDDVADDAEAAADAMD 563

QY 222 QDNVPVP--GKTPATAQSIQVDTATATQIEKGNAGIDAYFAGNAGSAGVANA----- 273

Db 564 LDSEMDIGEIVDVIADIDOTNNMLAL-----NASTEAARTGADGGFAVVADEVKTLA 617

QY 274 -KSNNSISNIDSAKAATATAKTQIAEAQKPPDPSPILOEAEQWVIOAEKDLKNIKPADGS 332

Db 618 EESRDAEDIESRLALQGVSDVADDEMRTSDT--VSDGRATVGDAAATLDDV----- 669

QY 333 DVPNPGTVGSKQGGSSIGSIRVSMLLDDAENETASILMSGFRQMIHMENTNPDSQAA 392

Db 670 -----VSFVADTDTAAGQIRAA--TDQHAASRVASAVDEV-----AGI 707

QY 393 QOELAAQARA-AKAGD-----DSAAALADAQAKALEAALGK 428

Db 708 SQETAQAQAVADSAAATQDTLSSVDDAAADLADRAALDDLLAE 752

RESULT 8

SPAP\_STRMU STANDARD; PRT; 1561 AA.

AC P23504;

DT 01-NOV-1991 (Rel. 20, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE CELL SURFACE ANTIGEN I/II PRECURSOR.

GN SPAP.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RC STRAIN-NG5 SEROTYPE C;

RA MEDLINE=90076473; PubMed=2687020;

RA Kelly C., Evans P., Bergmeier L., Lee S.F., Progulskie-Fox A.,

RA Harris A.C., Aitken A., Bleiweis A.S., Lehner T.;

RT "Sequence analysis of the cloned streptococcal cell surface antigen

RT I/II".

RL FEBS Lett. 258:127-132(1989).

RN [2]

RC STRAIN-NG5 SEROTYPE C;

RA MEDLINE=91207143; PubMed=1982405;

RA Kelly C., Evans P., Ma J.K.C., Bergmeier L.A., Taylor W., Brady L.J.,

RA Lee S.F., Bleiweis A.S., Lehner T.;

RT "Sequencing and characterization of the 185 kDa cell surface antigen

RT of Streptococcus mutans".

RL Arch. Oral Biol. 35:33S-38S(1990).

CC -!- FUNCTION: SURFACE PROTEIN ANTIGEN IMPLICATED IN DENTAL CARIES.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN; CELL WALL.

CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS

CC -!- IN THE REGION OF THE MEMBRANE ANCHOR.

CC -!- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: X17390; CAA35253.1; -

DR PTR: S06839; S06839; Gram\_pos\_anchor.

DR InterPro: IPR001899; Gram\_pos\_anchor.

DR Pfam: PF00746; Gram\_pos\_anchor; 1.

DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; 1.

KW Antigen; Signal; Transmembrane; Repeat; Cell wall; Dental caries.

FT SIGNAL 1 38

FT CHAIN 39 ? CELL SURFACE ANTIGEN I.

FT CHAIN 997 1561 CELL SURFACE ANTIGEN II (PROBABLE).

FT DOMAIN 39 1536 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1537 1556 MEMBRANE ANCHOR.

FT DOMAIN 1557 1561 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 60 550 HELICAL (POTENTIAL).

FT DOMAIN 219 464 3 X TANDEM REPEATS, ALA-RICH.

FT DOMAIN 847 963 3 X TANDEM REPEATS, PRO-RICH.

FT DOMAIN 1528 1533 CONSERVED IN GRAM-POSITIVE COCCI SURFACE

FT PROTEINS.

SEQ SEQUENCE 1561 AA; 170060 MW; 540D92768FC8AB4B CRC64;

Query Match 6.7%; Score 152; DB 1; Length 1561;

Best Local Similarity 22.3%; Pred. No. 1.2;

Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;

QY 48 KESKTDVSVERWSILRSVAVNALMSLADKLGIASSN-SSSSTSRSDVDSTTATATPTPPPT 106

Db 2 KVRKTYGFRKSKITLCGAVLGTVAASVAGQKVFADETTTSDVDVTKVVGQTGNPAT 61

QY 107 -----SDDYKTAQATAYDTTFTST-----SLADIAQALVS-LQDAVNTIKDT 147

Db 62 NLPEAQGSASKQAEQSQTCLERQWHTIEVPKTDLDQAADKASGVNVVDADVN-KGT 120

QY 148 AATDEETAAAEWE-----TKNADAIK-----VGAQITELAKYASNDQAILDSGLK 193

Db 121 VKTAAE-AVOKETEIKEDYTKQAEIDIKTTDOYKSDVAAHEAEVAKIKAKNQATKEQYK 179

QY 194 LTSFDLL-QTALLGVSANNKAAEL-----LKEMODNPVVPKTPALQAQSLVDQTDATAT 247

Db 180 ----DWAHKAEEVERINAANAASKTAYEAKLAQYQADLAQVKNANQASY-QKALAA 234

QY 248 QIE-----KDNAIGDAYFAGNAGSAGVENAKSNNSISNIDSAKAAIATAK-----TQIAEA 299

Db 235 QAEELKRVQEAANAATAAAY---DTAVAANNAKNTETIAAANEIEIRKRNATAKAEYETKLAQY 291

QY 300 Q---KKFPDPSPILOEAE-QWVIOA-EKDLKNIRPAD-----GSDVPNPGTVG 342

Db 292 QAEELKRVQEAANAATAAAYQAKLTAYOTELARVOKANADAKAAAYEAAVAAANNKNAALUTAE 351

QY 343 GS-----KQGGSSIGSIRVSMLLDDAENETASILMSGFRQMIHMENTE 385

Db 352 NTAIKQRNENAKATYEALQYQADLAQVKNANQASY-QKALAA 408

QY 386 NPDSQAQAEQELAAQARAATAA---GDDSAAL-ADAQAKALEAALGK 428

Db 409 NADAKAAVEAAVAAANAANAALTAENTAIAKRNADAKADYEAALAK 454

RESULT 9

HRPK\_PSESY STANDARD; PRT; 641 AA.

AC P41501;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE PATHOGENICITY LOCUS PROTEIN HRPK.

GN HRPK.

OS Pseudomonas syringae (pv. syringae).

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI\_TaxID=321;







```
QY 258 DAYFAGON-----ASGAVENAKSNNSISNIDSAKAAATATKTAQTAERAKK 302
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 241 -AYAAAEAYDKWARQVAAANAAAKAYEEAALAAANTAKN-DQIKAEIEAIQORSAGA--- 295
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 303 FPDSPILQEAROMVIOAEKDLKNIKPADGSDVPNPGTTVGSGKQGGSSIGSIRVSMLLDD 362
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 296 -----DYAKUAQOEKDLAAQAAGNAANEADYQAKKAAEQELARV-----Q 337
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 363 AENETASILMSGFRQIMHMFNTN-----PDSQAAQQLAAQARAAKAAGDDSAALALADA 418
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 338 AANAANK--QAYEQALAAANSKNAQITAENEALQON--AQAKA-----DYAKLAQY 385
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 419 QKALEAA-LGKAGQQQGLLNALG-----QIASAAVVSAGVLPLOQ 457
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 386 QKDLAAAQSGNAANEADYQEKLAAYEKELARVQAANAANKOAYEQVOQ 434
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 11
FLJB_SALTY
ID FLJB_SALTY STANDARD; PRT; 505 AA.
AC P52616; P97159;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PHASE-2 FLAGELLIN.
GN FLJB OR H2.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SL 375;
RX MEDLINE=95325331; PubMed=7541401;
RA Vanegas R.A., Joys T.M.;
RT "Molecular analyses of the phase-2 antigen complex 1,2,... of
  Salmonella spp.";
RL J. Bacteriol. 177:3863-3864(1995).
RN [2]
RP SEQUENCE OF 1-37 FROM N.A.
RX MEDLINE=82049491; PubMed=6271461;
RA Silverman M., Zieg J., Mandel G., Simon M.;
RT "Analysis of the functional components of the phase variation
  system.";
RL Cold Spring Harb. Symp. Quant. Biol. 45:17-26(1981).
RN [3]
RP SEQUENCE OF 482-505 FROM N.A.
RX STRAIN=SJ2353;
RA Mingorance J., Tanaka S., Tomimaga A., Enomoto M.;
RP Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
  FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -1- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
  BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
  PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
  or send an email to license@isb-sib.ch).
CC
CC EMBL; U1177; AAC43354.1;
CC EMBL; V01370; CAA24655.1;
CC EMBL; D26168; RAA05156.1;
CC ScyGene; SG10564; fljB.
CC InterPro; IPR001029; Flagellin.C.
CC InterPro; IPR001492; Flagellin.N.
CC Pfam; PF00700; Flagellin_C; 1.
CC Pfam; PF00669; Flagellin_N; 1.
```

```
DR PRINTS; PR00207; FLAGELLIN.
DR ProDom; PD000316; Flagellin_C; 1.
KW Flagella.
FT INIT-MET 0 0 BY SIMILARITY.
FT CONFLICT 37 37 I -> S (IN REF. 2).
SQ SEQUENCE 505 AA; 52404 MW; 3A5CC404AF7AF88B CRC64;

Query Match 6.5%; Score 148; DB 1; Length 505;
Best Local Similarity 21.0%; Pred. No. 0.5;
Matches 93; Conservative 59; Mismatches 168; Indels 122; Gaps 20;

QY 23 AQGLEASAAKSAEQAQRIAG-----AEAKPKESKTSVERWSILRSVNALMSLADKL 75
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 32 SSGLRINSKADDAAGQAIAANREFTANIKGLTQAASRNANDGISIAQTTEGALNEINNLRV 91
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 76 GIASSNSSSTSRSDVDSTTATPTPPPTSDDYKTAQ-----TAYDTIFT----- 123
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 92 RELAVQSANSTNSQSDLSISQAEI-TORLNEIDRVSGQTQFNGVKVLAQDNTLTIOYGAN 150
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 124 -----STSLADIAQAAVSLQ-----DAVTNKKDTAATD-----EETAIAAE 159
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 151 DGETIDIDLKQINSQTLGLDSLNVQKAYDKDTAVTTKAYANNCTTLDVSLDAAAKAA 210
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 160 WETKNADAIKVGQITELAKYASDQAILDSLGLKLTDFDLQ-----TALLQS 207
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 211 TGGTNGTASVTGGAV-----KFDADNNKYFVTIGGTCGADAANKGDYEVNVAATDGTVTLAA 266
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 208 VANNN-----KAELLKEMQDNPV-----PGKTPAIAGSLVDQTDATATQI-----EKD 252
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 267 GATKTTMPAGATTKTEVOELKDTFPAVVSADAKNALIAGG-VDATDANGAELVKMSYTDKN 325
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 253 GNAT-----GDAYFAG--QNASGAVENAKSN-----NSISNIDSAKAAI 289
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 326 GKIEGGYALKAGDKYAADYDEATGAIK-AKTSYTAADGTTTAANQLGGVD-GKTEV 383
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 290 AT-----AKTQIAEAKKPPDSPILQEAQEMVQIAE---KDLKNIKPA 329
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 384 VTIDGKTYNASKAAGHDFKAQPELAAAKTENP-LQKIDAAALQVDALRSLDGLAQNR 442
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 330 DGSVDVNPNGTIVGSGKQGGSSI 351
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 443 FNSAITNLGNTVNNLSEARSRI 464

RESULT 12
YJH8_YEAST
ID YJH8_YEAST STANDARD; PRT; 881 AA.
AC P47033;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 89.2 KDA PROTEIN IN SCP160-SMC3 INTERGENIC REGION.
GN YJL078C OR J1027.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX MEDLINE=96093911; PubMed=7483841;
RA Miosga T., Schaaff-Gerstenschlaeger I., Chalwatzis N., Baur A.,
  Boles E., Fournier C., Schmitt S., Velten C., Wilhelm N.,
  Zimmermann F.K.;
RT "Sequence analysis of a 33.1 kb fragment from the left arm of
  Saccharomyces cerevisiae chromosome X, including putative proteins
  with leucine zippers, a fungal Zn(II)-Cys6 binuclear cluster domain
  and a putative alpha 2-SCB-alpha 2 binding site.";
RL Yeast 11:681-689(1995).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
  INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
```



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; 249353; CAA8370.1; -  
 EMBL; X83502; CAA58492.1; -  
 EMBL; X88851; CAA61314.1; -  
 HSSP; P04284; ICFE.  
 SGD; S0003614; YJL078C.  
 InterPro; IPR001283; SCP.  
 Pfam; PF00188; SCP; 1.  
 PRINTS; P00837; V5TPXLIKE.  
 SMART; SM00198; SCP; 1.  
 PROSITE; PS01009; SCP\_AG5\_PRI\_SCT\_1; 1.  
 PROSITE; PS01010; SCP\_AG5\_PRI\_SCT\_2; 1.  
 Hypothetical protein.  
 DOMAIN 166 379 ALA/SER/THR-RICH.  
 SEQUENCE 881 AA; 89152 MW; 37DBAC660CA9D12A CRC64;

Query Match 6.5%; Score 148; DB 1; Length 881;  
 Best Local Similarity 21.8%; Pred. No. 0.99;  
 Matches 100; Conservative 73; Mismatches 199; Indels 86; Gaps 19;

QY 22 SAQGLEASAA--NKSAAQRTAGAEAKPKESKTSVERWSILRSVAVNALMSLADKLGTAS 79  
 DB 365 SAQTSSVSASLSVAADDIOGSTS--KEATSVSEHTSVTSATNA-AQYATRLGSSS 421  
 QY 80 SNSSSTSRSA-----DVSSTTATPTPPPTSDDYKTAQATYDTFT 123  
 DB 422 RSSSGAVSSAVSQSLNVLAVNTDYSVTSSTA-----HTTKDTATTSVTASEITS 476  
 QY 124 STSLADQALVSLQDAVNTNKDAATDETAIAAEWETKNADAIVKGAQITELAKYA-- 181  
 DB 477 ETAQAS-----SSTKEKINSNAATSSSI-----YSNSASVS-CHGVTYAAEVAIT 520  
 QY 182 -SDNAILDSLGLTSPDLLQALLOSVA--N--NKAELLKEMODNPVVPCKTPAIA 235  
 DB 521 SEQSSALATVPATNCSSIVATTLTENSSTTTITATKSTTLTANN-----STRAAT 575  
 QY 236 QSLVDQT-DATATQIEKDGNAGIDAYFAGQNASGAVENANNNSISNDSAKAAIATAKT 294  
 DB 576 AVTIQPTLDPTDNSAPTDNAKHTSTVGSSTGASLDSLRITTSISVSSNTTQLVST--- 632  
 QY 295 QIAEAQKFPDPSPILOEAQVIAQKDL-KNIPKADGSDVPNPGTVVGGSKQGGSSIGS 353  
 DB 633 --CTSESDYSPSPF--AISTATTESNLITNTITASCSTDSNFPPTSAASTDETAFTRT 688  
 QY 354 IRVS-MILDDAENETASTLMSGFQMIHMENT-----NTVVPASFPSTTTTCLENDDTAFSSIYTE 743  
 DB 689 ISTSCSLNGASTQTSETLTSPMKT-----NTVVPASFPSTTTTCLENDDTAFSSIYTE 743  
 QY 398 AOAARAAAGDSDSAAA-LADAQKALEAALGAKQOOG 434  
 DB 744 VNAATIINPGETSSLASDFATSEKPNPETSVMKSTSNEG 781

RESULT 13  
 HTR6\_HALSA STANDARD; PRT; 778 AA.  
 ID HTR6\_HALSA  
 AC Q48319;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HALOBACTERIAL TRANSUCER PROTEIN VI.  
 GN HTR4 OR HTPVI.  
 OS Halobacterium salinarium.  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.

OX NCBI\_TaxID=2242;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S9;  
 RX MEDLINE=96275896; PubMed=8674984;  
 RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,  
 RT Osterheilt D.;  
 RL FEMS Microbiol Lett. 139:161-188(1996).  
 CC -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL  
 CC TRANSDUCTION.  
 CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 6.5%; Score 147; DB 1; Length 778;  
 Best Local Similarity 22.9%; Pred. No. 0.96; Indels 58; Gaps 19;  
 Matches 97; Conservative 74; Mismatches 184;

QY 20 DLSAGGLEASAAKSAQAQRTAGAEAKPKESKTSVERWSILRSVAVNALMSL----- 71  
 DB 378 DAEEAARSEAAQAADAAQAQ--AEAEAAARESEAAQ-RRLETTAEAFSETMRYAAGDLTV 435  
 QY 72 ---ADKLGIASSN-SSSSTSRSDVSTTATPTPPPTSDDYKTAQATYDTFTTSTSL 127  
 DB 436 RLDAVFOQAAMADIAAFNEMAADMEATADAGR----FADEVATASTDASD-----SA 485  
 QY 128 ADIAQALVSLQDAVNTNKDAA--TDETAIAAEWETKNADAIVKGAQITELAKYASDQ 185  
 DB 486 AAVEQTGRVSDAVGRIFRDRAADQDQLEAVASETDEMSATIEEVAASADQVAE-TSORA 544  
 QY 186 AILDSLGLKTSFDLLQALLOSVA--N--NKAELLKEMODNPVVPCKTPAIAQTS--- 237  
 DB 545 AALGDGQAAQDAV--AQLEIEDETQAAATAVDQLEAKMSEIETIVAAITDIAEQTNM 602  
 QY 238 LVQDOTATATQIEKDGN---AIGDAY--FAGONASGAVE-----NAKSNNSISNDS 284  
 DB 603 LALNANIEAARADQDGGFVAVVAVEVDLADSKASAAEIALVAEVAETSVAMDR 662  
 QY 285 AKAAATATAKTOIAEAKKFPDPSPILOEAQVIAQKDL-KNIPKADGSDVPNPGTV 341  
 DB 663 IQERVSDGVETVSETERS-----LSEIAGRIAEADTGVQEISNAMDDQAAASVDVTTAV 716  
 QY 342 GGSQKQCGSSIGSIRVSMILLDDAENETASILMSGFQMIHMENTEN-PDSQAQQLAQA 400  
 DB 717 GDVAALGEETATEAESTA--DAAAEQAQSTLSDVAAQ-----TETLAHVALREHAAQF 768

401 RAA 403  
 769 EVA 771



```
Db 169 ASSGAEEAASAKATEAEKSAEESKNAATSAAGAAKTSET-----NAAASQOQAATSAAS 223
QY 146 DTAATDEETAIAAEWETKNADAIKVGAIQITELAKYASDNOAIILDSLGKLTSPFLLQTALL 205
Db 224 TAATKASEAATSA---RDVASKEAAKSETNASSAGRAASATAAENSARAATKS--- 277
QY 206 QSVANNKAAELKEMQDNVPVPGKTPALAQSLVDOTDATABATOLEKDGNAIGDAYFAGON 265
Db 278 ---ETNARSETAERSASAAADAKTAAGSASTATKAT-----EAAGSASVASQS 326
QY 266 ASGAVENA-KSNSISNIDSAKAAIA-----TAKTOIAEAKKPPSPILQEAEMVIOA 319
Db 327 KSAEAAAIRAKNSAKRAEDIAVALEADATTRKGIQVQLSSATSETLAATPKAVKV 386
QY 320 EKDLKNIK-PADG---SDVPNPGTTVGGSKQOQSSIGSIRVSMLLDDAENETASILMSGF 375
Db 387 VMDETNRKAPLDSPALGTGTPTAPTALRG-----NNTQTANTAFVIAAI 430
QY 376 ROMIHMENTENPDQSAQAQAAKAAAGDDSAALADAKKALEAALGRAGQOOGI 435
Db 431 ADVI---DASPDALNTLNELAA-----ALGND-----PDFATTMTNALAGKQPKNAT 474
QY 436 LNALQIATA 445
Db 475 LTALAGLSTA 484

RESULT 15
SYEP_DROME STANDARD; PRT; 1714 AA.
ID SYEP_DROME AC P28668; Q9VCF5;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BIFUNCTIONAL AMINOACYL-TRNA SYNTHETASE [INCLUDES: GLUTAMYL-TRNA
DE SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE); PROLYL-TRNA
GN AATS-GLOPRO OR CG5394.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92097547; PubMed=1756734;
RA Cerini C., Kerjan P., Astier M., Gratecos D., Mirande M., Semeriva M.;
RT "A component of the multisynthetase complex is a multifunctional
RT aminoacyl-tRNA synthetase."
RL ENBO J. 10:4267-4277(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=OREGON-R;
RC MEDLINE=97217441; PubMed=9063462;
RX Cerini C., Semeriva M., Gratecos D.;
RA "Evolution of the aminoacyl-tRNA synthetase family and the
RT organization of the Drosophila glutamyl-prolyl-tRNA synthetase gene.
RT Intron/exon structure of the gene, control of expression of the two
RT mRNAs, selective advantage of the multi-enzyme complex."
RL Eur. J. Biochem. 244:176-185(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agapayni A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
```

```
RESULT 14
STF_LAMB
ID STF_LAMB STANDARD; PRT; 774 AA.
AC P03764; P03745;
DT 21-JUL-1986 (Rel. 01, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SIDE TAIL FIBER PROTEIN.
GN STF.
OS Bacteriophage lambda.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA."
RL J. Mol. Biol. 162:729-773(1982).
RN [2]
RP IDENTIFICATION AS STF.
RX MEDLINE=92165720; PubMed=1531648;
RA Haggaard-Ljungquist E., Halling C., Calendar R.;
RT "DNA sequences of the tail fiber genes of bacteriophage P2: evidence
RT for horizontal transfer of tail fiber genes among unrelated
RT bacteriophages."
RL J. Bacteriol. 174:1462-1477(1992).
RN [3]
RP RECONSTRUCTION OF STF.
RX MEDLINE=93068310; PubMed=1439823;
RA Hendrix R.W., Duda R.L.;
RT "Bacteriophage lambda P2: not the mother of all lambda phages."
RL Science 258:1145-1148(1992).
CC -1- MISCELLANEOUS: THE COMMON LABORATORY STRAIN OF BACTERIOPHAGE
CC LAMBDA; LAMBDA PAPA; CARRIES A FRAMESHIFT MUTATION RELATIVE TO UR-
CC LAMBDA, THE ORIGINAL ISOLATE. THE UR-LAMBDA VIRIONS HAVE THIN,
CC JOINTED TAIL FIBERS (SIDE TAIL FIBERS) THAT ARE ABSENT FROM LAMBDA
CC WILD TYPE. RELATIVE TO LAMBDA PAPA, UR-LAMBDA HAS EXPANDED
CC RECEPTOR SPECIFICITY AND ADSORBS TO E. COLI CELLS MORE RAPIDLY.
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC CORRECTED IN POSITION TO 396 TO RECREATE THE ORIGINAL STF PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J02459; AAA96555.1; ALT_FRAME.
CC EMBL; J02459; AAA96557.1; ALT_FRAME.
CC PIR; A04389; QXBP1L.
CC PIR; A04370; QXBP2L.
CC InterPro; IPR000122; Chemotaxis_transducer.
CC InterPro; IPR001778; F0A_allergen.
CC Fiber protein.
CC SEQUENCE 774 AA; 77527 MW; CDD1DF85E919123B CRC64;

Query Match 6.58; Score 146.5; DB 1; Length 774;
Best Local Similarity 20.94; Pred. No. 1;
Matches 90; Conservative 66; Mismatches 209; Indels 65; Gaps 13;

QY 26 LEA5AANKSAQRIAGAEAKPKESKTDVSRVSWILRSVAVNALMSLADKLGIASSNSSSS 85
Db 110 VEEVARNASVVQASTADAKKSAGDASASAAQVAALVTDATDSARAATSAGQAAS-SAQE 168
QY 86 TSSRADVSDTATPTPPPTSDDYKTAQATPAYDTIFTSTSLADIQAALVSLQDAVNIK 145
```



RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Rao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zubing W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of *Drosophila melanogaster*.";  
RA Science 287:2185-2195(2000).  
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +  
CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU).  
CC -1- CATALYTIC ACTIVITY: ATP + L-PROLINE + TRNA(PRO) = AMP +  
CC PYROPHOSPHATE + L-PROLYL-TRNA(PRO).  
CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS  
CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE  
CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,  
CC ARGINYL, AND ASPARYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY  
CC PROTEINS, P18, P48 AND P43.  
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I  
CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II  
CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -1- SIMILARITY: CONTAINS 6 "WHEP-TRS" DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M74104; AAA28594.1; -  
CC DR Flybase: FBgn005674; Aats-glupro.  
CC DR InterPro: IPR002106; AA\_trna\_ligase\_II.  
CC DR InterPro: IPR000738; WHEP-TRS.  
CC DR InterPro: IPR000924; trna-synt\_1c.  
CC DR InterPro: IPR002314; trna-synt\_2b.  
CC DR InterPro: IPR004112; trna-synt\_1.  
CC DR InterPro: IPR002316; trna-synt\_1c; 1.  
CC Pfam: PF00749; trna-synt\_1c; 1.  
CC DR Pfam: PF00587; trna-synt\_2b; 1.  
CC DR Pfam: PF00458; WHEP-TRS; 6.  
CC DR PRINTS: PR00987; TRNASYNTHGLU.  
CC DR PRINTS: PR01046; TRNASYNTHPRO.  
CC DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_I; 1.  
CC DR PROSITE: PS00179; AA\_TRNA\_LIGASE\_II\_1; FALSE\_NEG.  
CC DR PROSITE: PS00339; AA\_TRNA\_LIGASE\_II\_2; FALSE\_NEG.  
CC DR PROSITE: PS00762; WHEP-TRS; 6.

KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Multifunctional enzyme; Repeat.  
FT DOMAIN 170 754 GLUTAMYL-TRNA SYNTHETASE.  
FT DOMAIN 755 800 WHEP-TRS 1.  
FT DOMAIN 827 872 WHEP-TRS 2.  
FT DOMAIN 901 946 WHEP-TRS 3.  
FT DOMAIN 980 1025 WHEP-TRS 4.  
FT DOMAIN 1055 1100 WHEP-TRS 5.  
FT DOMAIN 1129 1173 WHEP-TRS 6.  
FT DOMAIN 1174 1180 POLY-GLY.  
FT DOMAIN 1207 1714 PROLYL-TRNA SYNTHETASE.  
FT SITE 209 220 "HIGH" REGION.  
FT SITE 438 442 "KMSKS" REGION.  
FT BINDING 441 441 ATP (BY SIMILARITY).  
FT CONFLICT 102 106 TSPLP -> DKSIA (IN REF. 3).  
FT CONFLICT 233 234 VC -> AF (IN REF. 3).  
FT CONFLICT 341 345 NTACA -> KYCVR (IN REF. 3).  
FT CONFLICT 583 583 K -> R (IN REF. 3).  
FT CONFLICT 692 692 L -> A (IN REF. 3).  
FT CONFLICT 753 753 T -> S (IN REF. 3).  
FT CONFLICT 802 802 T -> S (IN REF. 3).  
FT CONFLICT 873 873 P -> T (IN REF. 3).  
FT CONFLICT 887 887 G -> V (IN REF. 3).  
FT CONFLICT 1201 1201 P -> PA (IN REF. 3).  
FT CONFLICT 1461 1461 MISSING (IN REF. 3).  
FT CONFLICT 1587 1587 G -> V (IN REF. 3).  
SQ SEQUENCE 1714 AA; 189197 MW; 6FE8C58045E48A8C CRC64;

Query Match 6.5%; Score 146; DB 1; Length 1714;  
Best Local Similarity 19.8%; Pred. No. 2.8;  
Matches 101; Conservative 74; Mismatches 192; Indels 142; Gaps 20;

Qy 7 PGPI-----DETERTPADLSAOGLEASAAKSAEQRIAG-----A 43  
Db 701 FSPVILFIPGHTKDVPTGLKVNAPDAKATKASSPVSSGGQASELDSQITQOGLVR 760  
Qy 44 EAKPKESKTDSVERWSILRSVAVNALMSLADKLGTASSSSSSSSRSADVDTSTAT-APT 102  
Db 761 DLKSKKAQKQID-----VAVKKLLAL-----KADYKSATGDKWKPQTSATSAVP 807  
Qy 103 PPTSDDDYKTAQATAYDTFTSTSLADIAQALVSLQDAVTNKTAADEETAIA----- 157  
Db 808 AASSS-----SANDAV-----SYNASIVKOGDLVRDLKGRKASKPEIDAIAVKTL 852  
Qy 158 ---AEWETKNADAIVKGAQITELAKYASDNOAI--LDSLGKLTSPDLLQTLALQSVANN 212  
Db 853 ELKAYKTLTGQDWKPGVPTAAPSASAAAPSVGNDVSAQILS-----QITAOGD 903  
Qy 213 KAAELLKEMQNPVVVPGKTPAIAOSLVDTATATQIEKDGNAIGDAYFAGONASGAVEN 272  
Db 904 KVRELKSAKADKATVDAAVKTLTLLSKADYKAATGSD-WKPGTT-----APAPAAAPKV 956  
Qy 273 AKSN--SISNIDSAKAATATAKTAEAKKPPDPILQEAQOMVIAEKDLKNI-- 326  
Db 957 KOEKNPDPASVLTNTLNKTAQOGDKIRQLKSAKSEKSEVLEAVKLLALKTDYKSLTG 1016  
Qy 327 ---RPADGSDVFNPGTT-----VGGSKQGGSGISGSI----- 354  
Db 1017 QEWKP--GTVAAPTPTVNVIDLTTGG--DSGSDVGSVLSKIAQOGDKIRKLKSEAKAKNVI 1072  
Qy 355 --RVSMIL-----DDAENETASILMSGFRQIMHMTENTENPQSQAAQOELA 397  
Db 1073 DPEVKTLLALKGEYKTLGKDWTPDAKSEPAVV-----KKEASPVSMASPAKDELTOEIN 1127  
Qy 398 AQAARAAGDSDSAAALADAKALEAAL 426  
Db 1128 AQGEVRAAKGNKAAKEVIDAEVAKLLAL 1156

Search completed: February 7, 2002, 21:42:45  
Job time: 599 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:41:15 ; Search time 172 Seconds  
(without alignments)  
393.745 Million cell updates/sec

Title: US-09-391-606-9

Perfect score: 2261

Sequence: 1 MNPFGPGIDETERTPPAD.....SAAVAGVLPQQLWLIRA 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL\_17.\*
- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2214	97.9	715	2 Q46166	Q46166 chlamydia p
2	2188	96.8	651	2 Q927H7	Q927H7 chlamydia p
3	451.5	20.0	647	2 Q84627	Q84627 chlamydia t
4	192	8.5	1327	2 Q9X7M2	Q9X7M2 staphylococ
5	192	8.5	2478	2 Q9RL69	Q9RL69 staphylococ
6	190	8.4	2478	2 Q9LCH2	Q9LCH2 staphylococ
7	182.5	8.1	2055	2 Q85472	Q85472 abiotrophia
8	182	8.0	2481	2 Q99QR6	Q99QR6 staphylococ
9	180	8.0	2271	2 Q99QX4	Q99QX4 staphylococ
10	177.5	7.9	1822	2 Q07290	Q07290 streptococ
11	177	7.8	697	5 Q9NDJ0	Q9NDJ0 plasmodium
12	174	7.7	1545	2 Q9BD01	Q9BD01 streptomyce
13	173	7.7	2016	5 Q9BIT0	Q9BIT0 plectreureys
14	171	7.6	627	1 Q9HS86	Q9HS86 halobacteri
15	171	7.6	993	2 Q99QZ5	Q99QZ5 staphylococ
16	171	7.6	1795	2 Q9LCJ9	Q9LCJ9 staphylococ
17	166.5	7.4	1128	5 Q26947	Q26947 trypanosoma
18	166.5	7.4	2232	5 P91365	P91365 caenorhabdi
19	164	7.3	6713	2 Q99U54	Q99U54 staphylococ

20	161.5	7.1	956	2 Q9LON7	Q9LON7 streptomyce
21	161.5	7.1	2045	2 Q9A0X5	Q9A0X5 streptococ
22	161.5	7.1	3381	2 Q9KX33	Q9KX33 streptococ
23	161	7.1	1038	10 Q9AS09	Q9AS09 oryza sativ
24	160.5	7.1	544	1 P71409	P71409 halobacteri
25	160.5	7.1	2178	2 Q9KWR3	Q9KWR3 streptococ
26	159	7.0	1579	11 Q9WMP1	Q9WMP1 mus musculu
27	159	7.0	1684	11 Q9WTQ5	Q9WTQ5 mus musculu
28	158.5	7.0	1156	2 Q9Z5A4	Q9Z5A4 streptomyce
29	158	7.0	1365	2 Q9Z5A4	Q9Z5A4 streptomyce
30	158	7.0	2186	2 Q99TB0	Q99TB0 mycoplasma
31	157	6.9	1344	2 Q9Z5A4	Q9Z5A4 mycoplasma
32	156.5	6.9	641	2 Q9KS57	Q9KS57 vibrio chol
33	156	6.9	990	13 Q91803	Q91803 xenopus lae
34	156	6.9	6677	5 Q9A435	Q9A435 caenorhabdi
35	155	6.9	845	5 Q9Y1P8	Q9Y1P8 plasmodium
36	155	6.9	892	3 Q9P3P5	Q9P3P5 neurospora
37	154.5	6.8	1262	5 Q20684	Q20684 caenorhabdi
38	154	6.8	1041	10 Q9ASA4	Q9ASA4 oryza sativ
39	154	6.8	1814	5 Q9B1S9	Q9B1S9 plectreureys
40	153.5	6.8	969	5 Q9NDI9	Q9NDI9 plasmodium
41	153.5	6.8	993	10 Q9ASL3	Q9ASL3 oryza sativ
42	153	6.8	1569	2 Q94183	Q94183 streptococ
43	153	6.8	2614	5 Q97054	Q97054 dictyosteli
44	152.5	6.7	653	2 Q9X6N1	Q9X6N1 rhizobium l
45	152.5	6.7	962	2 Q9RIE3	Q9RIE3 yersinia pe

## ALIGNMENTS

RESULT 1

Q46166 PRELIMINARY; PRT; 715 AA.  
AC Q46166;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
DE 76 KDA PROTEIN.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=94156481; PubMed=7509320;  
RX Perez-Melgosa M., Kuo C.C., Campbell L.;  
RT "Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76-kilodalton protein containing a species-specific epitope";  
RL Infect. Immun. 62:880-886(1994).  
DR EMBL; L23921; AAA23117.1;  
SQ SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;

Query Match 97.9%; Score 2214; DB 2; Length 715;  
Best Local Similarity 98.9%; Pred. NO. 3.9e-115;  
Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY	1	MVNPFGPGIDETERTPPADLSAOGLEASAAKSAEAOIAGAEAKPKESKTSVRSWSI	60
Db	257	LVNPFGPGIDETERTPPADLSAOGLEASAAKSAEAOIAGAEAKPKESKTSVRSWSI	316
QY	61	LRSAVNALMSLADKLGIASSSSSSSSRSADVDSSTATATPTPPPTSDYKTAQATYDT	120
Db	317	LRSAVNALMSLADKLGIASSSSSSSSRSADVDSSTATATPTPPPTSDYKTAQATYDT	376
QY	121	IFTSTSLADIQAALVSLQDAVNIKDTATDEETAAEWEETKNADAIKVGAIQITELAKY	180
Db	377	IFTSTSLADIQAALVSLQDAVNIKDTATDEETAAEWEETKNADAIKVGAIQITELAKY	436
QY	181	ASDNOAILDLSGLKTSFLLQALLOSVAANNKAAELKEMODNPVPGKTPAIAGSLVD	240
Db	437	ASDNOAILDLSGLKTSFLLQALLOSVAANNKAAELKEMODNPVPGKTPAIAGSLVD	496



Db	1	MVNP	IGGPI	DETF	TPPP	ADL	SAQGL	EAANK	SEBAQRI	AGAEAK	PKESK	TD	SVER	WSI	60																																													
Qy	61	LRSA	VALM	SLAD	KLG	LIAS	SNSS	SSST	SR	SAD	VD	ST	TAT	APT	PPPT	SD	YK	TO	AO	Y	TD	120																																						
Db	61	LRSA	VALM	SLAD	KLG	LIAS	SNSS	SSST	SR	SAD	VD	ST	TAT	APT	PPPT	FFD	YK	TO	AO	Y	TD	120																																						
Qy	121	IFT	ST	SL	ADI	QAAL	VS	LQ	DAV	NI	KD	TAT	DE	TA	IA	AE	WET	KN	DA	L	KV	GA	IT	EA	KL	AK	Y	180																																
Db	121	IFT	ST	SL	ADI	QAAL	VS	LQ	DAV	NI	KD	TAT	DE	TA	IA	AE	WET	KN	DA	L	KV	GA	IT	EA	KL	AK	Y	180																																
Qy	181	ASD	QA	IL	DS	LG	KL	T	S	F	DL	L	LO	T	AL	LQ	S	V	AN	NK	AE	LK	E	M	O	D	N	P	V	P	GK	T	PA	I	A	Q	S	L	V	D	240																			
Db	181	ASD	QA	IL	DS	LG	KL	T	S	F	DL	L	LO	T	AL	LQ	S	V	AN	NK	AE	LK	E	M	O	D	N	P	V	P	GK	T	PA	I	A	Q	S	L	V	D	240																			
Qy	241	Q	T	D	A	T	A	T	Q	I	E	K	D	N	A	I	G	D	A	I	F	A	G	O	N	A	S	G	A	V	E	N	A	K	S	N	S	I	S	N	I	D	S	A	K	A	A	I	A	T	A	K	T	O	I	A	E	A	Q	300
Db	241	Q	T	D	A	T	A	T	Q	I	E	K	D	N	A	I	G	D	A	I	F	A	G	O	N	A	S	G	A	V	E	N	A	K	S	N	S	I	S	N	I	D	S	A	K	A	A	I	A	T	A	K	T	O	I	A	E	A	Q	300
Qy	301	K	K	P	D	S	P	I	LQ	E	AQ	EW	I	Q	A	E	K	D	L	N	I	K	P	A	D	G	S	D	V	N	P	G	T	T	V	G	S	K	O	O	G	S	S	I	G	S	I	R	V	S	M	L	L	360						
Db	301	K	K	P	D	S	P	I	LQ	E	AQ	EW	I	Q	A	E	K	D	L	N	I	K	P	A	D	G	S	D	V	N	P	G	T	T	V	G	S	K	O	O	G	S	S	I	G	S	I	R	V	S	M	L	L	360						
Qy	361	D	D	A	E	N	E	T	A	S	I	L	M	S	G	F	R	O	M	I	H	M	F	N	T	E	N	P	D	S	O	A	Q	E	L	A	A	Q	A	R	A	K	A	G	D	D	S	A	A	A	A	L	A	D	A	Q	K	420		
Db	361	D	D	A	E	N	E	T	A	S	I	L	M	S	G	F	R	O	M	I	H	M	F	N	T	E	N	P	D	S	O	A	Q	E	L	A	A	Q	A	R	A	K	A	G	D	D	S	A	A	A	A	L	A	D	A	Q	K	420		
Qy	421	A	L	E	A	L	K	G	A	G	A	G	O	O	G	I	L	N	A	L	G	I	A	S	A	A	V	S	A	G	V	P	L	454																										
Db	421	A	L	E	A	L	K	G	A	G	A	G	O	O	G	I	L	N	A	L	G	I	A	S	A	A	V	S	A	G	V	P	L	454																										

```

RESULT      3
ID          084627 PRELIMINARY; PRT; 647 AA.
AC          084627;
DT          01-NOV-1998 (Tremblrel. 08, Created)
DD          01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT          01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE          CHLPN 76 KDA HOMOLOG.
GN          Cht622..
OS          Chlamydia trachomatis.
OC          Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX          NCBI_TaxId=813;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=D/UW-3/CX;
RX          MEDLINE=99000809; PubMed=9784136;
RA          Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.
RA          Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA          Davis R.W.;
RT          "Genome sequence of an obligate intracellular pathogen of humans:
RT          Chlamydia trachomatis.";
RL          Science 282:754-759(1998).
DR          EMBL: AE001333; AAC68226.1; .
KW          Complete proteome.
SQ          SEQUENCE 647 AA; 68525 MW; CODI4C2D74473625 CRC64;

Query Match           20.0%; Score 451.5; DB 2; Length 647;
Best Local Similarity 29.6%; Pred. No. 1.6e-17;
Matches 138; Conservative 85; Mismatches 192; Indels 51; Gaps

QY    2 VNPTGGPIDERTPPADLSAOGLEASAANKSAEQRIAGAEAKPKESKTSVERWSIL 61
       :||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    15 MNPIINGOI-----ASNSEKSEKESEA-----SPSASSSVSSWSFL 52
       : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    62 RSVAALMSLADKLGIASSNSSTSRSDVDYTATAPTPPPPTSDDY---KTQAQTAY 118
       : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    53 SSAAHALISLRD--AILKNKSPTDSLS-QLEASTSTS-TVTRVAARDYNEAKSNFDTAK 108
       : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    119 DTIFTSTSLADIQALVSLODAVTNIKDTAATDEETAIAAEWEIKNADAIVGAQILTALA 176
       : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    109 SGLENATTLAETKMDLMALQOMDERLAKOKAEVTRIKALQEKQEV----KLNLV 168
       : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```







```

Db 838 KMDAYNEVKQAATARKAQAATVSNATNEEVAEADAADAAQKGLHDIOVVKSKQEVADT 897
QY 88 RSADVDSSTA-TAPTPTPTPTSDDYKTAQATAYDT-----IFTSTSLADIQAALVSL--- 137
Db 898 KSKVLDKINAIOQAKVKPAAD---TEVENAYNTRKQEIQNSNASTTEEKQAAAYTELDTK 954
QY 138 -QDAVTNIKDTATD-----EETAIAAEWETKNADAIAKVGQAQITELAKYASDNQAILDS 190
Db 955 KQEARTNL-DAANTNSDVTTAKDNSIAAINOVQAAATTKKSDAK-AEIAQKASERKTAIEA 1012
QY 191 LGKLT-----SPDLLOTLALQSVAANNKAAELLKEMQDNPPVPGKT----- 231
Db 1013 MNDSTTEEQAAKDKVQAVVTANADIDNAA-----ANNVDVNAKTTNEATIAATPDA 1066
QY 232 ---PALAQSLVDOTDATATQIEKDGN-----ATGDYAFAGQNASGA 269
Db 1067 NVKPAKQAIAADKVOAQETAI--DGNNGSTTEEKAAAKQOVOTEKTTADAADAIDAAHTNAE 1124
QY 270 VENAKSNNSISNI-----DSAKAAIAT-----AKTOIAEAKKFPDPSPILOEA--E 313
Db 1125 VEAAK-KAAIAKTAIEAIQAPATTTKDNKAKEATKANERKTAIAQOTDITABEIAAANADVD 1183
QY 314 QMWIOAEKDLKNIKPADG-SDVPNPGTTVGSKQOQSSIGSIRVSMILLDDAENETASILM 372
Db 1184 NAVTOAN---SNIEAANSQNDVDQAKTTGENSIDVPTVYKKTAT-----ARNEITAILN 1235
QY 373 SGFRQMIHMFNTENPDPSQAAQELAAQARAAGAAGDDSA--ALADAQKALEAA 425
Db 1236 NKLEIQATPDATDEEKQAAAE--ANTENKANQAISSAATTNAQVDEAKANAEEAIAINAV 1293
QY 426 LGKAGOOQOGLNALGOI 442
Db 1294 TPVKVKKQAARDEIDQL 1310

```

```

RESULT 6
ID Q9LCH2 PRELIMINARY; PRT; 2478 AA.
AC Q9LCH2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE FMTB.
GN FMTB.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RA Komatsuzawa H.;
RT "Staphylococcus aureus gene for affecting the methicillin
RT resistance.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB025716; BAA93438.1;
SQ SEQUENCE 2478 AA; 262993 MW; 1C118EBE0DB03B34 CRC64;

```

```

Query Match 8.4%; Score 190; DB 2; Length 2478;
Best Local Similarity 23.9%; Pred. No. 0.026;
Matches 133; Conservative 85; Mismatches 195; Indels 144; Gaps 27;

QY 4 PIGPGPIE-----TERTPPADLSAQGLEASAANKSA-EAORIAGAEAKPKESKTDs--- 54
Db 780 PLNPOTTEVEAEALERINAKVS--GVKAIEATTTAOLERKNEEISKENITDSTQT 837
QY 55 -----VERWSILRSYVNALMS-----LADKLG-----IASSNSSSTS 87
Db 838 KMDAYNEVKQAATARKAQAATVSNATNEEVAEADAADAAQKGLHDIOVVKSKQEVADT 897
QY 88 RSADVDSSTA-TAPTPTPTPTSDDYKTAQATAYDT-----IFTSTSLADIQAALVSL--- 137

```

```

Db 898 KSKVLDKINAIOQAKVKPAAD---TEVENAYNTRKQEIQNSNASTTEEKQAAAYTELDTK 954
QY 138 -QDAVTNIKDTATD-----EETAIAAEWETKNADAIAKVGQAQITELAKYASDNQAILDS 190
Db 955 KQEARTNL-DAANTNSDVTTAKDNSIAAINOVQAAATTKKSDAK-AEIAQKASERKTAIEA 1012
QY 191 LGKLT-----SPDLLOTLALQSVAANNKAAELLKEMQDNPPVPGKT----- 231
Db 1013 MNDSTTEEQAAKDKVQAVVTANADIDNAA-----ANNVDVNAKTTNEATIAATPDA 1066
QY 232 ---PALAQSLVDOTDATATQIEKDGN-----AIGDAYFAGQNASGA 269
Db 1067 NVKPAKQAIAADKVOAQETAI--DGNNGSTTEEKAAAKQOVOTEKTTADAADAIDAAHTNAE 1124
QY 270 VENAKSNNSISNI-----DSAKAAIAT-----AKTOIAEAKKFPDPSPILOEA--E 313
Db 1125 VEAAK-KAAIAKTAIEAIQAPATTTKDNKAKEATKANERKTAIAQOTDITABEIAAANADVD 1183
QY 314 QMWIOAEKDLKNIKPADG-SDVPNPGTTVGSKQOQSSIGSIRVSMILLDDAENETASILM 372
Db 1184 NAVTOAN---SNIEAANSQNDVDQAKTTGENSIDVPTVYKKTAT-----ARNEITAILN 1235
QY 373 SGFRQMIHMFNTENPDPSQAAQELAAQARAAGAAGDDSA--ALADAQKALEAA 425
Db 1236 NKLEIQATPDATDEEKQAAAE--ANTENKANQAISSAATTNAQVDEAKANAEEAIAINAV 1293
QY 426 LGKAGOOQOGLNALGOI 442
Db 1294 TPVKVKKQAARDEIDQL 1310

```

```

RESULT 7
ID Q85472 PRELIMINARY; PRT; 2055 AA.
AC Q85472;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE EXTRACELLULAR MATRIX BINDING PROTEIN (FRAGMENT).
GN EMB.
OS Abiotrophia defectiva.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Abiotrophia.
OX NCBI_TaxID=46125;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NVS-47;
RA Manganello R., van de Rijn I.;
RT "Cloning and Characterization of emb, a Gene Encoding the Major
RT Adhesin of Streptococcus defectivus.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067776; AAD03320.1;
FT NON_TER 2055 2055
SQ SEQUENCE 2055 AA; 215640 MW; 9699C11DDDE93E2FD CRC64;

```

```

Query Match 8.1%; Score 182.5; DB 2; Length 2055;
Best Local Similarity 22.9%; Pred. No. 0.054;
Matches 109; Conservative 76; Mismatches 189; Indels 101; Gaps 21;

QY 19 ADLSAQGLEASAANKSAEQAQRIAGAE-----KPKESKTDsVERWSILRSYVNALMSLAD 73
Db 1371 ATKAKNAIDAATSNDETAKQNEGTQAINAVPQTPPKATD-----AKNAVTOAAD 1420
QY 74 KLGTASSNSSST-----SRSAVDSTTATPTPPPTSDDYKTAQATAYDTIFTSTSLA 128
Db 1421 RKKDAIENDPNLTREEKVAAKVDA-----EAKKAKDAIDAATNA 1462
QY 129 DIQAALVSLQDAVTNIKD--TAATDEETAIAAEWETKNADAIAKVGQAQITELAKYASDNQA 186
Db 1463 DVTAKQNEGTKAINDVPQTPPTAKTDAKNAVTOAADAKK-DAIEKDPNLTREEKDA--KA 1519

```



QY 187 ILDSLGKTSFLLQTALLQS--VANNKKAELLKEMQNDPNVVPKGTPTAIAQSLVDQTD 244  
Db 1520 KVDAEAKKAK-DAIDAATSNDVDTAKONEGTAKINDVPQTPT--AKTDA-KNAVQAADA 1575  
QY 245 TATOIEKDGNAIGDAYFA-----GONASGAVENAKSN-----NSISNI 282  
Db 1576 KDAIEKDPNLTREKDKAKAKVDAAEKAKDAIDAATSNDVDTAKDAGKNAINAVPQT 1635  
QY 283 DSAKAAIATAKTOIAEAOQ-KFPDSPILQEAQVIOAEKDLKLNKPKADGSDVPNPGTIV 341  
Db 1636 PTAKTDAKNAVQAADAKKDALENDANLTREKDAKAKVDAAEKAKNAIDAATSNDV 1695  
QY 342 GSKQOQS-SIGSI-RVSMLLDDAEN---ETASILMSGFQMIHFNTE-----PSQ 390  
Db 1696 TAKONEGTAKINDVPQTPTAKTDAKNAVDOATDKKSAI-----ENDPALFREKD 1746  
QY 391 AAQQLAAQAAKAAKAGDDSAALADAAQKALEAALGKAGQOQILNALGOITASA 445  
Db 1747 AKAKVDAAETKAKNAIDAATSNDVDTAKQ-----DAG--KNAINAVPQTPTA 1792

## RESULT 8

Q99QR6 PRELIMINARY; PRT; 2481 AA.  
AC Q99QR6;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DE FMTB PROTEIN.  
GN FMTB(MBP) OR SA1964.  
OS Staphylococcus aureus subsp. aureus N315.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=158879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,  
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,  
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,  
RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,  
RA Yanashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,  
RA Ogasawara N., Hayashi H., Hiramatsu K.,  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus.";  
RL Lancet 357:1225-1240(2001).  
DR EMBL; AP0031136; BAB43253.1; -.  
KW Complete proteome.  
SQ SEQUENCE 2481 AA; 263767 MW; E1EAB99B81665E8 CRC64;

Query Match 8.0%; Score 182; DB 2; Length 2481;  
Best Local Similarity 23.3%; Pred. No. 0.072;  
Matches 128; Conservative 91; Mismatches 202; Indels 128; Gaps 25;  
QY 4 PIGPGPIDE-----TERTPPADLSAQGLEASAANKSA-EAORIAEAKPKESKTD--- 54  
Db 763 PLNPDTTNEEVAELERINAKVS--GVKATERTTADLERKNEEIKFNIENITDSTQT 820  
QY 55 -----VERMSILRSVNALMSLADKLGIASSNSSSSTSRAD-----VDSTTATAPT 101  
Db 821 KMDAYKEVQAATARKAQNATVSNATDEEVAEANAADAAOEGLEHDIQVYKSOQEVADT 880  
QY 102 PP-----PPTSDDYKTOAQATYDT-----IFTSTSLADIQNALVSL--- 137  
Db 881 KAKVLDKINAIOTQAKVPAAD---TEVENAYNTRKQIONSNASTTEKEAAYTELDK 937  
QY 138 -QDVAYNIKDTAATD-----EETAIAAEWETKNADAIVKGAQITELAKYASDNOAIDLS 190  
Db 938 KQEAETNL-DRAANTNSDVTYAKNGIAAINQVQAATTKSDAK-AEIAQKASERTATIEA 995  
QY 191 LKGLT-----SFDLLQALQSVANNKA-----AELKEMQNDPNVVPKGT--PAI 234

Db 996 MNDSTTEEQAQAKQVQAVVTANADIDNATANTVDNAKTNEATIAAITPDANVPAA 1055  
QY 235 AQSLVDQTDATATOIEKDGNA-----IGDAYFAGQNASGAVENAKSN 277  
Db 1056 KOAIADKVOAQETALDANGSTTEKEAQQVQTEKTAADAIDAHSNVEVEAK-NA 1114  
QY 278 SISNI-----DSKAAIAT---AKTOIAEAOQKFPDSPILQEA--EOMVIOAEK 321  
Db 1115 EIAETAIQPAVTTTKDNKQAIAITKANERKTAIAQTQDITAEIEAANADVDNAVTOAN- 1173  
QY 322 DLKNIKPADG-SDVPNPGTGVGSKQOQSSIGSIRVSMLLDDAENETASILMSGFQMIH 380  
Db 1174 --SNIEAANSQNDVQAKTGTGETSIDVTPVTNKKAT-----ARNEITAILNKKLOEQ 1226  
QY 381 MFENTPDSQAAQQLAAQAAKAAKAGDDSAAL-----ALADAQKALEAALGKAGQO 433  
Db 1227 TPDATDEEKQADAE--ANTENKANKAISAATTAQVDEAKANAEEAALNAVTPKVVKKO 1284  
QY 434 GILNALQOI 442  
Db 1285 AAKDEIDQL 1293

## RESULT 9

Q99QY4 PRELIMINARY; PRT; 2271 AA.  
AC Q99QY4;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DE SA2447 PROTEIN.  
GN SA2447.  
OS Staphylococcus aureus subsp. aureus N315.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=158879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,  
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,  
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,  
RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,  
RA Yanashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,  
RA Ogasawara N., Hayashi H., Hiramatsu K.,  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus.";  
RL Lancet 357:1225-1240(2001).  
DR EMBL; AP0031136; BAB43752.1; -.  
KW Complete proteome.  
SQ SEQUENCE 2271 AA; 227847 MW; 7C2A7040D6C8289D CRC64;

Query Match 8.0%; Score 180; DB 2; Length 2271;  
Best Local Similarity 18.9%; Pred. No. 0.084;  
Matches 79; Conservative 96; Mismatches 187; Indels 56; Gaps 12;  
QY 22 SAQGLEASAANKSAPAQRIAGAEAKPKESKTDSEVERWSILRSVNALMSLADKLGIASSN 81  
Db 1516 SASASESSSSSTSLDSTSSASQSSSDSTSSASLSLSTSTSNRMT-----IASLS 1570  
QY 82 SSSSTSRSDVDSSTTATAPTPTPPPTSDDYKTOAQI-ADYTTFTSTSLADIQNALVSLQDA 140  
Db 1571 TSVTSSESGSTSESDSTSTSDSTSSRSTSSASGSTSTSTSDSRSTASTS-- 1628  
QY 141 VTNIKDTAATDETAIAAEWETKNADAIVKGAQITELAKYASDNOAIDLSGLKLSFDL- 199  
Db 1629 -TSMRTSTSDQSMSLSLSTSTSSMSDS-----TSLSDSVSDST--DSTASTSGSMS 1678  
QY 200 LOTALQSVANNKAELLKEMQNDPNVVPKGTPTAIAQSLVDQTDATATOIEKDNAGDA 259  
Db 1679 VSLSDSTSTSTSEVMSASISD-----SQSMSESVNDSSESVSESDSKSMGS 1732



QY 260 YFAGONASGAV-----ENAKSNNSISNIDSAKAIATATAKQTAEAKKPPDSPILQEA 312  
 Db 1733 TSVSDSGSLSVSTSLRKESVSESSLSGSGMSDSVSTSDSSLSVSTSLRSESVSSES 1792  
 QY 313 EQMVIQAEKDLNKPAGDPVNPPTVGGSKQGGSSIG-----SIRVSMILLDDA 363  
 Db 1793 D-----SLSDSKSTSGS-----TSTSTSGSLSTSTSLSGSESVSESTSLSDSISMSDS 1840  
 QY 364 ENETASILMSGFROMIHMFNTENPDSQAQOELAAQARAACAAGDDSAALADAKA 421  
 Db 1841 TSTSDSLSGSLSGSTSLSTSDLSKSLSS-----SQSMGSESTSTSVSDSQSS 1895  
 RESULT 10  
 Q07290 PRELIMINARY; PRT; 1822 AA.  
 AC Q07290;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE EF PROTEIN.  
 GN EPP\*.  
 OS Streptococcus suis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1307;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1890;  
 RX MEDLINE=93328288; PubMed=8333363;  
 RA Smith H.E., Reek F.H., Vecht U., Gielkens A.L.J., Smits M.A.;  
 RT "Repeats in an extracellular protein of weakly pathogenic strains of  
 RT Streptococcus suis type 2 are absent in pathogenic strains.";  
 RL Infect. Immun. 61:3318-3326(1993).  
 DR EMBL; X71880; CAA50714.1;  
 SQ SEQUENCE 1822 AA; 192631 MW; 3838960C77641D7D CRC64;

Query Match 7.98; Score 177.5; DB 2; Length 1822;  
 Best Local Similarity 23.68; Pred. No. 0.088;  
 Matches 116; Conservative 77; Mismatches 195; Indels 103; Gaps 21;  
 QY 13 TERTPPADLSAQGLEASAAKSAEARIAGAEAKPKES-KTDSVVERMSIL-RSAVNALMS 70  
 Db 1201 TEKAAAEELAGEKSLDTGKEARDAVELAKDKELAKEAIRTEEEAEATKIVEKLAEDTRKA 1260  
 QY 71 LADKLGIASSNSSSTSSAD-VSTTATATPTPPPTSDDYKTQATAYDTFTSTSLAD 129  
 Db 1261 IEDNPNLSDEKQAEIKKLTDAVAKTLATI-----RDNADKRTQEAERKA-----QALAD 1309  
 QY 130 IQAALVSLQDAVTNIDKTAATDEETAIA--AEWETKNADAIAKVGAGITELAKYASDNOAI 187  
 Db 1310 LEKA-----KETQKTADKAIDRLTILYKDGLEATKQDAKNKIADKAAAEKATASPN 1364  
 QY 188 LDSLGLKLSFDLQALQSVANNKA-----AELKEMQDNPVPGKTPAIQAQSLVD 240  
 Db 1365 LTDAEKKFTD-----AVDAEAKANDAIASAATSPADVQKE-EDAGVA-----ATAEDVLD 1414  
 QY 241 QTDATA-TQIEKDGNAGIDAYFAGONASGA-----VENAKSNNSISNIDSAKAI 289  
 Db 1415 AAKQDAKNKIADKAAAEKATGNSPNLTDAEKKTTFDVADEAVAKANDAIS----- 1465  
 QY 290 ATAKTQIAEAKKFPDSPILQAEQWVIAEKDLNKPAGDSV-----PNPQGT--- 340  
 Db 1466 --AATSPADVQKE-EDAGVAATAEDVLDAAKQDAKN-KIAKESDAKSAIDANPNLTDAE 1521  
 QY 341 -----VGGSKQGGSSIGSIRVSM-----LDDAENETASILM 372  
 Db 1522 KESAKKAVDADAKAATDAIDASTSPVEAQSAEDKGVGSIADVDLAAKQDAKNKIADKAA 1581  
 QY 373 SGRFQMIHMFNTENPDSQAQOELAAQARAACAAGDDSAALADAKAALAAAGKAGQ 432  
 Db 1582 AAKEAIDANPNLSDAEKAASKAVDADAKATTDAID--ASTSPVEAQSAEDKGVGSI--R 1637

QY 433 QGILNALGOIA 443  
 Db 1638 QDVLDAKQDA 1648  
 RESULT 11  
 Q9NDJ0 PRELIMINARY; PRT; 697 AA.  
 ID Q9NDJ0;  
 AC Q9NDJ0;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE MEROZGITE SURFACE PROTEIN 3B.  
 GN MSP-3B.  
 OS Plasmodium vivax (strain Belem).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=31273;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BELEM;  
 RA Galinski M.R., Al-Khedery B., Ingravallo P., Corredor-Medina C.,  
 RA Barnwell J.W.;  
 RT "Plasmodium vivax merozoite surface proteins-3 beta and -3 gamma share  
 RT structural similarities with Plasmodium vivax MSP-3 alpha and define a  
 RT new gene family in Plasmodium.";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF099662; AAF78287.1;  
 DR InterPro: IPR000122; Chemotaxis\_transducer.  
 SQ SEQUENCE 697 AA; 75685 MW; 0AF2AE9801A958A7 CRC64;

Query Match 7.88; Score 177; DB 5; Length 697;  
 Best Local Similarity 20.98; Pred. No. 0.029;  
 Matches 118; Conservative 74; Mismatches 186; Indels 186; Gaps 19;  
 QY 1 MVPNIGPGPIDETERTPPA-----DLSAGGLEASAAKSAEAKQRI 40  
 Db 53 IVNPDGSDVTDDEGGEALGQSGPEKSAEPKVAHQAEVNEKSKAQNAKAEKA 112  
 QY 41 AGAEAKPKESKTDSDVERWSI-----LRSAVN 66  
 Db 113 AKAAESAKKNTDLAEKVNPVPTLNNYKFAESAATAEQKOENIATAEKKVAANGEV 172  
 QY 67 ALMSLADKLGIASSNS-----SSSTSRSAVDSTTA----- 97  
 Db 173 ELQKLKDEVDKAAKAKQQLKAEIAEHAVKAKVAKTEAEKAKQDATTAEKAIKETGTS 232  
 QY 98 -----TAPTPPPPTSDDYKTQATAYDTFTSTSLA-----DIAAALVSLQDAVT 142  
 Db 233 KSENVTKAIDMAKKEEETKNQASIASENADKAAKAAQAEVKEIKDENKEISQLENEIT 292  
 QY 143 NIKTAAATDETAIAEWETKNADAIAKVAQI-----TELAKYASDNOALDLSIGLITS 196  
 Db 293 KLGILNVTVKELASNAEDASNAKKEKWKVAQIAAEVAKAEKAEKAEANFLAEKAK--- 349  
 QY 197 FDLQTLALLQSVANNKAAELLKEMQDNPVVPKTPAIQAQSLVDOTDATAFTQIEKDGNAI 256  
 Db 350 ---QTA-EKIAKTSKSTEKITE-----EVRRATEFAKTAGDATTQAATE-----AA 391  
 QY 257 GDVAFAGONASGAVENAKSNNSISNIDSAKAIATAKTO-----IAEAKKFPDSPILQE-- 311  
 Db 392 GDVSESEKQKQVLESIK-QKAESALQASKDAI-KAKTEAENFLEIAKEVKEPAEAKAEEA 449  
 QY 312 -----AQMVIAEKDLNKPAGDSVNPPTVGGSKQGGSSIGSIRVSMILLDDAEN 365  
 Db 450 QKATSADEAKTEALKIAEEVYKSDAS-----EN 478  
 QY 366 ETATILMSGFRQMIHMFNTE-NPDSQAQOELAAQARAACAAGDDSAALADAKAALAE 424  
 Db 479 EKKKI-----ETEANATAGEAKAAFAKAAKADAKDNTNEAVTTLAFAKEVKEK 526  
 QY 425 ALGRAGQOQGGILNALGQTASAANV 448







```
Db 1689 AAAASASS 1696
RESULT 14
Q9HS86 PRELIMINARY; PRT; 627 AA.
AC Q9HS86;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HTR14 TRANSDUCER.
GN HTR14 OR VNG0355G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OC NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
FX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE004994; AAG18922.1;
DR InterPro; IPR000122; Chemotaxis_transducer.
DR InterPro; IPR000658; DUF5.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00672; DUF5; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
KW Complete proteome.
SQ SEQUENCE 627 AA; 65655 MW; E5F2867FA6CA2A75 CRC64;

Query Match 7.6%; Score 171; DB 1; Length 627;
Best Local Similarity 23.4%; Pred. No. 0.054;
Matches 120; Conservative 87; Mismatches 184; Indels 122; Gaps 27;

QY 17 PPADLSAQGLEASANKSEAEQRIAGAEAKPK-----ESKTSVSRWS-ILRSVAVNMLSL 71
DB 145 PALDESVPVGFGEISITEMADSLSEAYTALEEDKTALEHQQAELERQSEOLRALVDALSEA 204
QY 72 ADKLGIASSNSSSTSRSDVDSTTATPTPPPTSDDYKTQA-----QTAYDTIFTS 124
DB 205 TD-----AARAGDL---TATVDAAALDYDDHRAAEVDFENQLLEATDISI 249
QY 125 TSLADIQAAL-----VSLQDAVNIKDTA--ATDETAIAAEWETKNADA 167
DB 250 QSFSDAVLAVSRITDERVDAVDRSAVSEVTEIADGANQOTNOLNIAAEMDTVSATV 309
QY 168 IKVCAQITELAKYASDQAIL---DSLGLK-TSPDLLQALLOSANNKAAELLKEMO 222
DB 310 EETASANDVAKTA---QAAADGEDGRGEVEETIETALRALREQSQVAETVESLAAEVE 366
QY 223 DNPWFVPGKTPAIAQSLVDQDTATA--TQIE-----KDGNAIGDAYFAGQASGAVENAKS 275
DB 367 R--IDGIT-ALIEDIAEETNMLALNASTEAARTSGDGD--GFVAVVADVEKDLAEETREQ 420
QY 276 NNSISNIDSA-----KAATATAKTQIAEAQKFFPSP-ILQEAQEVQIAEKDLKNKP 328
DB 421 AADISEIVDAVTEKAEDASTAIGEVD-AEVEKITKAEGVLDRFEAIV-----DEVANVH 475
QY 329 A--DGSVDPNPG-----TTVGGSKQOGSSI-----GSRVSMLLDD 362
DB 476 AVQEISATDQGAQSYTDVVGWVEEYASVSEETAESDTVAADNAEQTDATDEVDQMD 535

Query Match 7.6%; Score 171; DB 2; Length 993;
Best Local Similarity 24.8%; Pred. No. 0.095;
Matches 102; Conservative 62; Mismatches 151; Indels 96; Gaps 20;

QY 76 GIAS---SNSSSTSRSDVDSTTA---TAPTPTPTSDDYKTQAQATYDTIFTTSLA 128
DB 366 GVASTPPSAPSGDTAPNNNTQNTAPNSNAPVSTTPOSTSGKKGOSFADITTOVSTA 425
QY 129 D-----LQAAVSLQDAVNIKDT---AATDETA-----IAAEWET-- 162
DB 426 NENTQNTIDKDVKSNEAALTGSLLSLNNLDTQAKAAQKQSQALRNISYGLASDKFSDF 485
QY 163 -KNADAIKVGAQITELAKYASDQAILDSLGLK---TSFDLLQALLOSANN--NKAAE 216
DB 486 RESLDNVKSGLEYT--TQY---NQOFIDTLEIKENNVNDSKEIDKVKAAANNRINSLR 540
QY 217 ILKEMODNPVPGKT-PAIAQSLVDQ---TDATATQ-----LEKQDNA-----IGDAY 260
DB 541 LVNQI--SNALKNGSSGTAEATKLLDQLSKLSSLSSEFRDYKVKDLNSSLVSISQIMDEL 599
QY 261 FAGQASGAVENAKSNNSISNIDSAKAAIAATKTAIAEAQKFFDPSPILQEAQEVQIAE 320
DB 600 NKGOTALSINVQ--SKLNTIDOVINSQAILKNCKTFIDRLQ-----TVLPSTEQQYISAV 652
QY 321 KDLKNIKPADGSDVPNPGTTVGGSKQOGSSISGSIKVSMLLDDAENETASI-----MSG 374
DB 653 KNAQANFPKVSIDVAKAANFVRNDLPQ-----LEQRLTNATASVKNKLPFLTLNG 701
QY 375 FROMIHMTENTPDSQAQOELAAQARAAGDDSAALADAOKALEAA 425
DB 702 YDQAVGLNKNQPOAKKALSILA-----DFSQNKLPDVEKDLKA 741
```



Fri Feb 8 08:45:30 2002

us-09-391-606-9.rspt

Page 9

Search completed: February 7, 2002, 21:41:19  
Job time: 4648 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:33:50 ; Search time 12230.8 Seconds  
(without alignments)  
1923.418 Million cell updates/sec

Title: US-09-391-606-12  
Perfect score: 1426  
Sequence: 1 tggacgtactctgtgtgc.....tccttgacctggaagtgc 1426

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8246589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_om.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1098	77.0	1170	1	AF131889	AF131889 Chlamydia
2	1098	77.0	1170	15	AE033787	AE033787 Chlamydia p
3	1098	77.0	1830	1	CHTMOMPP	M69230 Chlamydia p
4	1098	77.0	20057	1	AE001652	AE001652 Chlamydia
5	1098	77.0	25150	1	AE002167	AE002167 Chlamydia
6	1098	77.0	300550	1	AP002547	AP002547 Chlamydia
7	1098	77.0	1170	1	CHTMOMPB	M6064 Chlamydia p
8	1066	74.8	1602	1	CPMOMPX	X72023 Chlamydia p
9	997.2	69.9	1170	1	CHTMOMPEQ	L04982 Chlamydia p
10	967	67.8	999	1	CHTMOMPAI	M73038 Chlamydia p
11	668.4	46.6	670	1	AF131230	AF131230 Chlamydia
12	665.2	46.6	670	1	AF131229	AF131229 Chlamydia
13	606.6	42.5	1425	1	AF269259	AF269259 Chlamydia
14	600.2	42.1	1170	1	AF272945	AF272945 Chlamydia
15	600.2	42.1	1261	1	CHTMOMPAAD	M73036 Chlamydia p
16	600.2	42.1	1337	1	CPOMP	X51859 Chlamydia p
17	600.2	42.1	1338	1	CHTMOMPAAAA	L39020 Chlamydia p
18	600.2	42.1	1435	1	AF269256	AF269256 Chlamydia
19	595.8	41.8	1436	1	AF269267	AF269267 Chlamydia
20	592.6	41.6	1430	1	AF269266	AF269266 Chlamydia
21	591.8	41.5	1392	1	AF269279	AF269279 Chlamydia p
22	585.8	41.1	1176	1	CHTMOMPAH	L25436 Chlamydia p
23	585.8	41.1	1412	1	AF269260	AF269260 Chlamydia
24	585.8	41.1	1421	1	AF269261	AF269261 Chlamydia
25	581.2	40.8	1482	1	AF269282	AF269282 Chlamydia
26	574.2	40.3	1387	1	AF269280	AF269280 Chlamydia
27	573.4	40.2	1200	1	AF269257	AF269257 Chlamydia
28	573.4	40.2	1200	1	AF269258	AF269258 Chlamydia
29	573.4	40.2	1227	1	CHTMOMPAH	AF269257 Chlamydia
30	571.8	40.1	1343	1	CPMPNOMP	M73037 Chlamydia p
31	555.6	39.0	1434	1	AF269269	X61096 Chlamydia p
32	552.2	38.7	1107	1	CPDNAX	AF269269 Chlamydia
33	543	38.1	1024	1	CHTMOMPAK	Z18756 Chlamydia p
34	536	37.6	1058	1	CPAJ5617	M73040 Chlamydia p
35	534.4	37.5	1058	1	CPAJ5613	AJ005617 Chlamydia
36	534.4	37.5	1058	1	CPAJ5614	AJ005613 Chlamydia
37	532.8	37.4	1058	1	CPAJ5615	AJ005614 Chlamydia
38	532.8	37.4	1058	1	CPAJ5618	AJ005615 Chlamydia
39	532.8	37.4	1058	1	CPJ004875	AJ005618 Chlamydia
40	529.6	37.1	1445	1	AF269265	AJ004875 Chlamydia
41	529.6	37.1	1610	1	CHTMOMPXX	L04980 Chlamydia p
42	528	37.0	1058	1	CPJ004874	AJ004874 Chlamydia
43	528	37.0	1393	1	AF269262	AF269262 Chlamydia
44	528	37.0	1432	1	AF269264	AF269264 Chlamydia
45	528	37.0	1438	1	AF269263	AF269263 Chlamydia

## ALIGNMENTS

RESULT	1	DNA	BCT	31-MAR-1999
AF131889	AF131889	Chlamydia pneumoniae major outer membrane protein (MOMP) gene, complete cds.		
LOCUS	AF131889	AF131889.1 GI:4545320		
DEFINITION	AF131889	Chlamydia pneumoniae		
ACCESSION	AF131889	Chlamydia pneumoniae		
VERSION	AF131889	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.		
KEYWORDS	AF131889	1 (bases 1 to 1170)		
SOURCE	AF131889	Sriram,S., Mitchell,W.M. and Stratton,C.W.		
ORGANISM	AF131889	Multiple sclerosis associated with Chlamydia pneumoniae infection of the CNS		
REFERENCE	AF131889	Neurology 50 (2), 571-572 (1998)		
AUTHORS	AF131889	98145402		
TITLE	AF131889	2 (bases 1 to 1170)		
JOURNAL	AF131889	Mitchell,W.M., Tharp,A.C., Stratton,C.W. and Sriram,S.		
MEDLINE	AF131889	Direct Submission		
REFERENCE	AF131889			



JOURNAL Submitted (26-FEB-1999) Pathology, Vanderbilt University, C-3321  
 Medical Center North, Nashville, TN 37232, USA  
 FEATURES Location/Qualifiers  
 source 1. .1170  
     organism="Chlamydotheca pneumoniae"  
     /db\_xref="taxon:83558"  
     note="from the cerebral spinal fluid of a patient with  
     rapidly progressive multiple sclerosis"  
 gene 1. .1170  
     /gene="MOMP"  
 CDS 1. .1170  
     /gene="MOMP"  
     /codon\_start=1  
     /transl\_table=11  
     product="major outer membrane protein"  
     /protein\_id="AAD22492.1"  
     /db\_xref="GI:454321"  
     /translation="MKLLKSAALLSAAFAGSVLQALPVGNPSDPLSLIDGTIWEGA  
     AGDPDPCATWCDALSLRAGFYDVDFRILKVDAPKTFSGMAKPTGSAANLYTTAVD  
     RPNATYKHLHDAEFNTAGFIALINDFDFCTLCASNGYIRGNSTAFNLVGLFGV  
     KGTIVANALEPNVLSUNGVGVLLITDTSFSDVGARGALWECGCATLGAERFOYAKSPK  
     VELNLYCNVDSFVSNVKNPKYGVAFPLTDAGVATATGKATINLTHMQVGSASLY  
     RLMSLPYIGVQMSRATFDADRIATPKPLTAVLNLTAWNPISLGNATLALSTDDSES  
     DFMQIVSQINKFKSRKACGVTVGATLVDAADKWSLTAEARLINERAHHVSGQFRF"  
 variation 54  
     /gene="MOMP"  
     /note="silent mutation"  
     126  
     /replace="c"  
 variation 126  
     /gene="MOMP"  
     /note="silent mutation"  
     /replace="g"  
 BASE COUNT 320 a 254 c 246 g 350 t  
 ORIGIN  
 Query Match 77.0%; Score 1098; DB 1; Length 1170;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-294;  
 Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 129 ttgcctgttaggaaccccttgatccaaaccttataattgattgatacattatgggaagt 188  
 Db 70 TTGCCGTGATAGGAAACCCCTTCATGCCAAGCTTATTATGATGGTACATATGGGAAGT 129  
 QY 189 gctgcaggagatccttgcgactccttgcgacttcttggtgcgacgtattagcttaoigt 248  
 Db 130 GCTGCAGGAGATCCTTGGAGCTCTTGCGCTACTTGGTGCAGCGTATTAGCTTACGTGT 189  
 QY 249 ggaattacggagactatgttttcgacgatactcttaaaagtagatgcactaaacattt 308  
 Db 190 GGATTTTACGGAGACTATGTTTCGACCGTATCTTAAAGTAGATGACCCCTAAACATT 249  
 QY 309 tctatggagcaagcactgactggatccgctgctgctcaactatactactgcgtagataga 368  
 Db 250 TCATGGGAGACCAACGCTACTTGGATCCGCTGCTCAACATATACTGCGGTAGATAGA 309  
 QY 369 cctaaccggcctacaataagcatttacacgatacgagtggttccactaagcaggtctc 428  
 Db 310 CCTAACCCGGCCTACAAATAAGCATTTACACGATCGAGTGGTTTCACATAAGCAGCTTC 369  
 QY 429 attgccttaaacatttggagatcgcttgaatttcttgcactcttagagcttctaagt 488  
 Db 370 ATTGCCCTTAAACATTTGGGATCGCTTGAATTTCTGTACTTTAGGAGCTTCTAATGGT 429  
 QY 489 tacattagaggaaactctacagcttcaatcgttggtttatttcggagtgtaaggtact 548  
 Db 430 TACATTAGAGAAACTCTACAGCGTTCATCTCGTTGGTTTATTTCGGAGTAAAGGTACT 489  
 QY 549 actgtaaatgcaatagaactacaaacattcttctaagtaacgagtggttgaaacttac 608  
 Db 490 ACTGTAATGCAANTGAATACCAACGCTTCTTTAAGTACGAGGTGTTGACTTTAC 549  
 QY 609 acagacacctcttctcttggagcgttaggcgctctggagcctctatgggaatgcggtt 668

[illegible]







/db\_xref="GI:144541"  
/translation="MKRLKALLSAAAFAGSVGLQALPVGNPSPSLIDGTIWECA  
AGDPDPCATWCDATSLRAGFYGVDFDRILKVDAPKTFSGAKPTGSAANYTTAVD  
RPNPYNKHLHDAEFTNAGFIALNIWDFDFTCLGASNGYIRGNSAFNLVGLFGV  
KGTIVNKLHDAEFTNAGFIALNIWDFDFTCLGASNGYIRGNSAFNLVGLFGV  
VEELNVCNVSQFVSNKPKYKGVAFPLPDAGVATGATKFSATINHEWQVGSLSY  
RLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNITAWNPSSLGNALSTTDSFS  
DFMOIVSQINKEKSRKACGVTGATLVDAADKWSLTAEARLINERAHVSGQFRF"  
386..1483  
mat\_peptide /note="putative"  
stem\_loop 517 a 383 c 355 g 575 t  
BASE COUNT 1502..1528  
ORIGIN

Query Match 77.0%; Score 1098; DB 1; Length 1830;  
Best Local Similarity 100.08; Pred. No. 4.2e-294;  
Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 129 ttgcctgtagggaacccctctgtatccaaagcttatttaattgattggtgacgctatttagcttgcgt 188  
DB 386 TTGCCTGTAGGGAACCCCTTCTGATCCAAAGCTTATTAAATGATGGTACAAATATGGGAAGGT 445  
QY 189 gctgcagagatcccttgatccctgctactctgtgctgacgctatttagcttgcgt 248  
DB 446 GCTGCAGAGATCCCTTGCCTGCTTCTGCGTACTTGTGCGGAGCTATTAGCTTACGTGCT 505  
QY 249 ggattttacggagactattttccacgcgtatcttaaaagttagatgcacctaataacatt 308  
DB 506 GGATTTTACGGAGACTATGTTTTCGACCGTATCTTAAAGTAGATGCACCTAAACATTT 565  
QY 309 tctatgggagcaagcactgactgactgctgctgcaactactactgctccttagataga 368  
DB 566 TCTATGGGAGCCAAAGCTACTGGATCCGCTGCTGCAAACTATACACTCCCGTAGATAGA 625  
QY 369 cctaaacccgctcacaataagcatttacacgagcagagtggttccactaatcagcgttc 428  
DB 626 CCTAACCCGCTTACAAATAGCATTTACAGGATGCGAGTGTTTCACTAATCAGGCTTC 685  
QY 429 attgccttaaacatttgggagcgttctgtatgtttctgtacttttagagccttcaatgt 488  
DB 686 ATTGCCTTAAACATTTGGGATCGCTTGTATGTTTCTGTACTTTAGAGCTTCTAATGCT 745  
QY 489 tacattagaggaactctacagcgttcaactctctgttctgttattcggagtttaaggtact 548  
DB 746 TACATTAGAGGAACCTCTACAGCGTTCACTCCTGTTGTTTATTCGAGCTTAAAGTACT 805  
QY 549 actgtaaatgcaaatgaactaccacacgcttcttcttaagtaacgaggttctgaactttac 608  
DB 806 ACTGTAATGCAAAATGAACCTACCAACCGTTCTTTTAAAGTAACGAGCTTCTGAACCTTAC 865  
QY 609 acagacacctcttctcttgagcgttagcgcctctgtgagccttatgggaatcggtgt 668  
DB 866 ACAGACACCTCTTCTCTTGAGCGTAGGCGCTCTGAGCGCTTATGGGAATCGGCTGT 925  
QY 669 gcaactttggagctgaattcccaatgcaacagcagctcccaacacctaataagtaagcaactaa 728  
DB 926 GCAACTTTGGAGCTGAATCCCAATATGCACAGCTCCCAACCTAAAGTTGAAGAACTTAAT 985  
QY 729 gtgactgttaacgtatgcgaattctctgttaacaaacacacacacacacacacacacacac 788  
DB 986 GTGATCTGTACGTATCGCAATTTCTCTGTAAACAAACCCCAAGCGCTTATAAGCGCTGCT 1045  
QY 789 ttccctctgcaacagacgctggcgttagcagcagctactggaacaaagctctgcgacacac 848  
DB 1046 TTCCCTTTGCCACAGACGCTGGGTAGCAACAGCTACTGGAACAAAGTCTGCGACCATC 1105  
QY 849 aattatcatgaatggcaagtaggagcctctctatcttcaacagacataactcttagtgcca 908  
DB 1106 AATTATCATGAATGGCAAGTAGGAGCGCTCTATCTTACAGACTAAACTCTTTAGTGCCA 1165  
QY 909 tacattggagtagcattgctcagcgaacttttgatgctgataacatcccgactgctcag 968

DB 1166 TACATTGGAGTACAAATGGTCTCGAGCAACATTTTGTATGCTGATAACATCCGATTCAG 1225  
QY 969 ccaaaactactacagctgttttaacttaacttaactgcagtggaaccccttcttactaggaaat 1028  
DB 1226 CCAAACTACTACAGCTGTTTAACTTAACCTGCAATGGAACCCCTTCTTTTACTAGGAAAT 1285  
QY 1029 gccacagcattgctactactgctgcttctcagacttcacaaattgtttcctcgtcag 1088  
DB 1286 GCCACAGCTTGTCTACTACTGATTCGTTCTCAGACTTCATGCNAATGTTTCTCCTGICAG 1345  
QY 1089 atcaacaagttaaactagaaagcgttggaggttacttagagcagctacttagttgat 1148  
DB 1346 ATCAACAGTTTAACTAGAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT 1405  
QY 1149 gctgataaattgctcacttactgacgaagcgtctttaaacttaacagagagcgtgcacgta 1208  
DB 1406 GCTGATAAATGGTCACTTACTGCAAGAGCTGTTTAACTAACGAGAGAGCTGCTCACGTA 1465  
QY 1209 tctgctcagttcagattc 1226  
DB 1466 TCTGTCAGTTCAGATT 1483  
RESULT 4  
AE001652 20057 bp DNA BCT 01-DEC-2000  
LOCUS Chlamydia pneumoniae section 68 of 103 of the complete genome.  
DEFINITION AE001652 AE001363  
ACCESSION AE001652.1 GI:4376997  
VERSION  
KEYWORDS  
SOURCE Chlamydia pneumoniae CWL029.  
ORGANISM Chlamydia pneumoniae CWL029  
REFERENCE 1 (bases 1 to 20057)  
AUTHORS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE 2 (bases 1 to 20057)  
AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
TITLE Direct Submission  
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
FEATURES  
source 1..20057  
/organism="Chlamydia pneumoniae CWL029"  
/strain="CWL029"  
/db\_xref="taxon:115713"  
339..1508  
/gene="ompA"  
/note="Cpn0695"  
339..1508  
/gene="ompA"  
/codon\_start=1  
/transl\_table=11  
/product="Major Outer Membrane Protein"  
/protein\_id="AAD18834.1"  
/db\_xref="GI:4376998"  
/translation="MKLLKALLSAAAFAGSVGLQALPVGNPSPSLIDGTIWECA  
AGDPDPCATWCDATSLRAGFYGVDFDRILKVDAPKTFSGAKPTGSAANYTTAVD  
RPNPYNKHLHDAEFTNAGFIALNIWDFDFTCLGASNGYIRGNSAFNLVGLFGV  
KGTIVNKLHDAEFTNAGFIALNIWDFDFTCLGASNGYIRGNSAFNLVGLFGV  
VEELNVCNVSQFVSNKPKYKGVAFPLPDAGVATGATKFSATINHEWQVGSLSY  
RLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNITAWNPSSLGNALSTTDSFS  
DFMOIVSQINKEKSRKACGVTGATLVDAADKWSLTAEARLINERAHVSGQFRF"  
complement(173..1803)  
/product="rRNA-Gly"  
1892..2725  
/gene="rs2"  
/note="Cpn0696"



```

CDS
1892..2725
/gene="rs2"
/codon_start=1
/transl_table=11
/product="s2 Ribosomal Protein"
/protein_id="AAD18835.1"
/db_xref="GI:4376999"
/translation="MESQCKLTIKDLSAGAHGHOTRRNPKMKLYIFEEKNGLYI
INLAKTLQALNPHIRKVIQDNKTVLFVGTKKQAKCVIREAAIEAGEFFIAERWLG
GLMTNTTIRSIKTLDKIEDLSRNQYLTKEAALLAKRHOKLRLNLEGIRYMKKA
PGLLVVDPSPYEKIAVAEAKKLGIPVALVDTNCDPTPIDHVPCNDDSLKIRLIIN
VKENIEAKHKLGIETVSPYKSLVPPDLSAFESSQDDDESDEENREEDLLAKKFGEA
N"
2725..3573
/gene="tsf"
/note="Cp0697"
2725..3573
/gene="tsf"
/codon_start=1
/transl_table=11
/product="Elongation Factor TS"
/protein_id="AAD18836.1"
/db_xref="GI:4377000"
/translation="MSDFSMTLTKLROQTGVLTKCKEALACGNGLEAAVYLRKL
GLASAGKEHRTKEGIIAAKTDAAGTALIEVNETDFVANNVAFREVFNNLLDILK
YKVDTEALSAASQSPSLSVDELRAVTMTQVGENIRISVAYFPKATNSTVGIYSH
NGKTVALTMLSGSTADSLAKDIAMHVVAQOPLSKESVPAEAIKAEKEVIAQIQ
GKQPEVIEKIVTGKLTFFQEAQLLEQPFIKNADLSIQSLIDDFSKTSGSSVAIEQFI
LWKICA"
3581..4327
/gene="pyrH"
/note="Cp0698"
3581..4327
/gene="pyrH"
/codon_start=1
/transl_table=11
/product="UMP Kinase"
/protein_id="AAD18837.1"
/db_xref="GI:4377001"
/translation="NAKQTRRVLFKISGEALSKDSNRIDEMRLVSELRAVRNND
IFIALVGGNLTGLAEQKLOINRVADQMGLATLINGMAVADAKAEIDPCLLT
STLSCPOLADLYTPOKSIEALDOGKILICTCAGSPYLTTDGAALRACELNVDVLIK
ATMHDVGYDDPRFPDPAKYDFVSKDFSLNQLGVMDASALSICMDSHIPRVSF
LOHSLKALFDPTIGTLVSEDVNHVCSPRH"
4305..4847
/gene="trf"
/note="Cp0699"
4305..4847
/gene="trf"
/codon_start=1
/transl_table=11
/product="Ribosome Releasing Factor"
/protein_id="AAD18838.1"
/db_xref="GI:4377002"
/translation="MSVLODTEKKMAALDFHKEVKFSFRCKAHPALVETVVVDYV
TWRSLDIASISVADRLQVISPYDGNASAIKAGIIAANLNLQPEVSGSIIRKVE
FPADYQEMIKOLRKEAEAKINVRIRREANDKLKDSALTEDVYVGNKKIQELTD
KFCQKDELTKKEAEIASI"
complement(4945..5019)
/product="tRNA-Glu"
complement(5045..5117)
/product="tRNA-Lys"
5220..5735
/gene="Cp0700"
5220..5735
/gene="Cp0700"
/codon_start=1
/transl_table=11
/product="Cp676 hypothetical protein"
/protein_id="AAD18839.1"
/db_xref="GI:4377003"
/translation="MVHSPTHOCYHCQOPATICYTEIDKDKVIRSYVCATCPSPSHY
NNEHLSKSGVGLTLEGNCNCKTVVHRSKQDDQLLGCHQCITYTNFKNQITSLKSERV
SSFTMERGGGSLHIGRAPGEASNTNPLLLKLIALNEALQDTLEREDYEQAAVIRQDIN
HLTKNPDPS"
5722..6798
/gene="karg"
/note="Cp0701"
5722..6798
/gene="karg"
/codon_start=1
/transl_table=11
/product="Arginine Kinase"
/protein_id="AAD18840.1"
/db_xref="GI:4377004"
/translation="MTLPNDLLETLVKKESQANKVMPVTTFSIARNLSVSKFLPCL
SKEQLEITQITTSFHNIEGFEFIVLPKDTPLMKFEFLLEHFLDYLVPNGPEGE
ALVVSRSQDGLAAINFQDHLVIGDFQGNVEKTDQLQVLDSDYHSLKSLFAFSEFG
FLTTPNKGCTGLKSOCLHPIPALYSKEFTNLIDEEVEITTSLLGVGTGPPGNIVV
LSNRCSLGLTRELLSLRITASKLSVAEAAKRLSEENSQDLKNLLRSLGLLTHS
CQLEKETLDALSWIQLDIGLGIKAVTENHPLWNPFLWQIRAHALQKQAEQSDLDQ
KDTISHLRASVLKELTKGLSPESF"
complement(6831..6903)
/product="tRNA-Thr"
complement(7049..9808)
/gene="yscC"
/note="Cp0702"
complement(7049..9808)
/gene="yscC"
/codon_start=1
/transl_table=11
/product="probable Yop proteins translocation protein
C/general secretion pathway protein"
/protein_id="AAD18841.1"
/db_xref="GI:4377005"
/translation="MKTVILNIGRKILQIIRKKKKIKIGLSGLFFLDLVLVSSQRP
TETSANVKNLRDEKLAACPKNASALSAAKSKTKTTPGSTPSKVSFSDATQDKTF
QKTSQSAFPKPTTLKELEKKPRPERTADVRSPRFLPTQEVPRVPAASKEQL
DSIQWEEKQNTARRAVANALNISIKKLEEQSTVTEKDVQPKQTQATPHASKNVASP
QFVSKISGTFNFDSDQFNVTIVSHDPTSYDDLSITLLQVKMHLKLVVGGNNVL
IYRNPHLSKLSVTVDSSLTCEAVVTVRFLYSVSPSAVNIQPLSHDAIVSA
SEADRHVIVISDIAGNVKVDLLAALDCGTSVDMTEYEVKYANFAALVSYQDVLGT
LAEDDAFOMFTQPNKIFVYSSPLANKAEQLKSLDVPMAHTLDLPASPALGG
TGTSFSLRFMYKLYKQNGEVIANALQDIGNLYVTTAMDEDFINTLNSIQWLEVN
NSVIATIGQGNVDRVIGLNGLDLPKKQYIEVLIDTSLSKSWDFGQWVALGDEQS
KIVYASGLNNNTGIATPTKATVPPTPNPSPITPLTQGLTQFSDMLNSSAFGLGII
GNVLSHGKGSFLTGLGSLADQDGTIVLNPRIQAQDQQAQSFVGTQPTVTKT
IIQETGTVQIDYEDIGVNLVVTSTVAPNNVTLQIEQITSELHSAQSGLTPVTKT
YAATRLQIPDGCFLVMSGHIDKTKTVYSGVPLNLSIPLIRGLFSRTIDQOKRNMIM
ETPKVVISSEFEGTVTKNGEYRWAEDEGSMQVAPRAPECQPPSLQAESDFKII
EIAQ"
complement(9805..11313)
/gene="Cp0703"
complement(9805..11313)
/gene="Cp0703"
/codon_start=1
/transl_table=11
/product="S/T Protein Kinase"
/protein_id="AAD18842.1"
/db_xref="GI:4377006"
/translation="MDCRGGIPLPEPOVIGVHVKKILSKLRSRVVHGLHPETRHST
VIKVPSPSPSTSRVYNFLKQAQSLHOITHNIVKFRYCKWQDCLYIAMEYIEGIS
LREYILAQFISLPQADIIIDIAQALEHLHSRNLHDKPENILKIPGKTKLIDFG
LADWTEIQRAPHSVIGTPYNSPQSGESHPASDIYALGGLAYELTLGHLISLGRV
FLSLVPERISKILAKALQPPNNRYSSSTREFIQDIIHHYRSGMDQEDIRIKDHTVALY
BOLOQTFWLAPLTFPFDISGVLYHQGYPLYPHAYDTLLGEGDVNLWLGYSPI SNA
TIALSVKSLVCOODLQRLPLDRVCEINEELRMKIPIDEMGISILCLIEISKENKELS
WTACGTVFWIKRQGVQDPEFSGLGKLTSLQIRETKVAMETGDAVCTLELE
SVASLKTLSLAEQLQDRRAKAFPCPIESIHGGIQSRQSGNSPSTLISLKRIR"
complement(11329..12444)
/gene="flin"
/note="Cp0704"
complement(11329..12444)
/gene="flin"
/codon_start=1

```



```

/transl_table=11
/product="Flagellar Motor Switch Domain/YscQ family"
/protein_id="AADI1843.1"
/db_xref="GI:4377007"
/translation="MAVAADSSASMKGRNNPFLSGKTEQVAAPFPELCOHKIR
FKPRLFDVQVSIKFRGIVATAEATFQVGHLLIPMVQVPEVENLFLFLSEEDLQEL

```

	Query Match	77.0%	Score 1098;	DB 1;	Length 20057;
	Best local similarity	100.0%;	Pred. No. 5.5e-294;		
	Matches 1098;	Conservative	0;	Mismatches 0;	Indels 0;
Qy	129	ttgcgttaggaaccctctgaccgaagcttattatggtgacaaatggaaggt	188		
Db	408	TTGCTGTAGGAACCTTCTGATCCAAGCTTATTAAATGATGCTACAATATGGAAGGT	467		
Qy	189	gctcaggagatccttgcgataccttgcgctacttgggtgcgacgctattagcttcgct	248		
Db	468	GCTCAGGAGATCCTTGGCATCTTGGCTACTTGGTGGCAGCTATTAGCTTACGTGCT	527		
Qy	249	ggatttacgagactatgtttcgacgctatcttaaaagttagatgcacctaaacattt	308		
Db	528	GGATTTTACGAGACTATGTTTTCACCGTATCTTAAAGTAGATGCACCTTAAACATTT	587		
Qy	309	tctatggagccaagcctactggatccgcgtgctgcaaaactatactcgcgtagataga	368		
Db	588	TCTATGGAGCCAAGCCTACTTGGATCCGCTGCTGCAAACTACTACTGCCGTAGATAGA	647		
Qy	369	cctaacccggcctacaataaacatttacacatgacaagtggttcactaatgcaggttc	428		
Db	648	CCTAACCCGGCCTACAATTAAGCATTTACAGATGAGAGTGGTTACATTAATGCAGCTTC	707		
Qy	429	attgccttaaacatttggatcgcgtttgatgttttctgactcttaggagctcttaaggt	488		
Db	708	ATTGCCTTAAACATTTGGGATCGCTTTCATGTTTCTGACTTTAGGAGCTTCTAATGGT	767		
Qy	489	tacatagagaaactctacagcttcaactctggttttatccggaattaaagtaact	548		
Db	768	TACATTAGAGAAACTCTACAGCTTCAATCTGTTGTTTATTCGGAGTTAAAGGTACT	827		
Qy	549	actgtaaatgcaatgaaactaccaaactttctttaagtaacggagttgttgaactttac	608		
Db	828	ACTGTAAATGCAATGAACACTACCAACGTTTCTTTAAGTAACGGAGTTGTTGAACTTTAC	887		
Qy	609	acagacaccttcttcttggagcgtgagcgtctgagccttatgggaatgcggttct	668		
Db	888	ACAGACACCTCTTCTCTTGGAGCTAGCGCTCGTGGAGCCTTATGGSAATTCGCGTGT	947		
Qy	669	gcaactttggagctgaattccaatatcacagctccaaacttaaagttgaagaactaat	728		
Db	948	GCAACTTTGGGAGCTGAATTCCAATATGCACAGTCCAAACCTTAAAGTTGAAGAACTTAA	1007		
Qy	729	gtgatctgtaacgatacgcaattctctgtaaacaaaccaagggctataaagcgctgt	788		
Db	1008	GTGATCTGAAGTATCGCAATTCCTGTAAACAAACCAAGGGCTATAAAGCGCTTGCT	1067		
Qy	789	ttccctcttccaagaacgctggcgtagcaacagctacttggaacaaagctctgcgaccatc	848		
Db	1068	TTCCCTCTGCCAACAGACGCTGGCGTAGCAACGCTACTTGGAAACAAAGTCTCGACCATC	1127		
Qy	849	aattatcatgaatgcaagttaggacctctctatcttacagactaaactcttagtgcca	908		
Db	1128	AATTATCATGAATGGCAAGTAGAGGCTCTCTATCTTACAGACTAAACTCTTTAGTGCCA	1187		
Qy	909	tacattggagtaacaatggcttcgagcaacttttgatgctgataaaccgcgacttgctcag	968		
Db	1188	TACATTTGGAGTACAATGGTCTCGAGCAACTTTTGTATGCTGATAAACATCGCATTTGCTCAG	1247		
Qy	969	ccaaaactactacagctgttttaaaacttaactgcatggaacaccttcttactaggaat	1028		
Db	1248	CCAAAACTACCTACAGCTGTTTTAAACTTAACGTGATGGAACCCCTTCTTTACTTAGGAAAT	1307		
Qy	1029	gccacagactgttctactactgattcgttctcagacttcatgcaaatgttttctcctcag	1088		



```

CDS
/gene="CP0033"
1692..2084
/gene="CP0033"
/translation="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37928.1"
/db_xref="GI:7188973"
/translation="MLEKLIKNFATYMGITITLLELDADGAVLPISEVVKVRAQONAD
NEIVLSASIGALPSADYAKLYLQMMIGNLDFRGRETGGALGDSEGNVVMVRRFSGDT
TYDDFVRHVESFMFSETWLSDLGLGQ"
2098..4635
/gene="CP0034"
2098..4635
/gene="CP0034"
/translation="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37929.1"
/db_xref="GI:7188974"
/translation="MAVRLIVDEGLSGVIFVLEDIGISIGROSSANDIPIEDPKLG
ASQAIINTKDGSIYITNLDITPIVWNGVAIQETTLKNEDTILLGNOYKFLSDPEF
PDQIVYDFDIPEENFSNDGILSDSNEOGKLEPROTSETNHSKPKKELKFDQSSD
PITSGDOLADAFLASAKAEKNOPRAKYAKGLKLESSNESLNPKQONAKDSPKEERT
NKPQNAIMEDNGASPRQDPQPKSAEPLSKNTARDETPLEKNKPVEEKAKKATDPSPE
KKDQPEEGSKGSKIATPLDSQKESDEKAEAEFVQEEENETEDNKEDSDSAADA
NDDTASDHADNETPKPVNEKSAVLSFPHVQDLFRDQTIIPAEIDDIHAKNISV
DLTQSPRELLKVLAGANIGAEFHLDGKTYILGTDPTCDIVFNDLSVSHQAKITVG
NDGGTLEIDLSKNGVIVEGRKIDKSTLSSNOVVAGCTTLFLLDHAPADTIVASL
SPDDYLFGRQDAERLEORAEKOKRATLPAGSFILTLFVGGILLEGITAS
LEHTKEVPLNIDYQDLAQVINGFPVRYTPFNKNSQLFLIGHVKNSTDKSELLYK
VDALSVKSDNVNIDDEAVQWEMNILLISKRPEFGISMHSPEPKFIITGYVTEEQ
AACLYDLNIHFNLSENKVVYETOMLKAIALHQQGFSANIHAFVFNVEVILTYQ
VNNDARFRAVVOELSGIPGVRLVKNFAVLLPAREGIDILNLRYPNRYRVYGSRYG
EISINVVNGRILLTRGDVIDGMTVTSIQPNAIFLEKEGLKYKIDYNK"
4682..4930
/gene="CP0035"
4682..4930
/gene="CP0035"
/translation="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37930.1"
/db_xref="GI:7188975"
/translation="MFNMNTAKEKNQSPLLDLEQDMQDHDRAQELKASVQDKVHKL
HALLREGSDKESFGQQSLLAGYVALQKVLGRINKMI"
4952..5206
/gene="CP0036"
4952..5206
/gene="CP0036"
/translation="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37931.1"
/db_xref="GI:7188976"
/translation="MATNKSCTAFDNKMLDGVCTYVKGVOQYLTELETSTQGTVDLG
TMFNLFQRMQILSQYMESVSNLLTAVNTEMITMARAVKS"
5225..5674
/gene="CP0037"
5225..5674
/gene="CP0037"
/translation="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11

gene
/product="conserved hypothetical protein"
/protein_id="AAF37932.1"
/db_xref="GI:7188977"
/translation="MADLEVFQADFALLFEAGLIAIKOGDEDSARKKLPQSULHILNPNH
YCHDLGLALISLHKMDFDAERLSALIKGNEKNSKAFSLTHMLJLVHOGSSFEV
RRSLESLKFAQDOVIANCKIESTRALAQSVLDNHDITLVAKSAGPLG"
5708..6382
/gene="CP0038"
5708..6382
/gene="CP0038"
/translation="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37933.1"
/db_xref="GI:7188978"
/translation="MIDPVECFPNLDGDAEQSITQNSGTPLASELKKDIPFALGSGY
AAPKDTTLVQCFKPNMAMQONSLDPELOALESELEQEOINNLKGRLWDFRST
FEDSOTTAQFADEHFOAVGVIIIDLINEDLNTTAEHTQODARKDEKESGVTKKIIDWV
SSGEVLNALLYFSDRGNRESLANFLKVQYAVQVQRAELFASIVGTSSVSKTI
MITQLG"
6384..7712
/gene="CP0039"
6384..7712
/gene="CP0039"
/translation="type III secretion cytoplasmic ATPase ScTN"
/translation="similar to SP:P21211 GB:U02499 SP:P21212 SP:P40290
PID:437202; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="type III secretion cytoplasmic ATPase ScTN"
/protein_id="AAF37934.1"
/db_xref="GI:7188979"
/translation="MDQLITDFILMSQLGDVNLTVVGRITVEVGMILKAVVPNVRV
GEVCLVRNGMEPLTVVVGTFQSFAPLUGELSGVSPSSSEVPTGLPLHFRAGNGL
LGRVNLGEPIDVETKGLQNVQTFPIFRAPDPLHRAKLQILSTGVRICIDGMLT
VARQRIQIFAGAGYKSLGMIARNAEADVNVIALIGRGRVRFIEGDLGEEG
MKRSVIVYSTDSQSLNAAVYTAETAEYFROGKTVLWMLMDSVTFRAFALREVGL
AAGPPPARAGTTPSVFSTPLRLERSGASDKGTITAFYTVLVAGDMNPEVADEKSI
LDGHVLSNALAQAYHPAIDVLASISRLLPVPEQRRIIGKAREVLAKYKANEML
IRIGEYRGRSDREIDFALDHIDKLNRFUKQDIHEKNTYEAAQQLRAIFR"
7730..8236
/gene="CP0040"
7730..8236
/gene="CP0040"
/translation="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37935.1"
/db_xref="GI:7188980"
/translation="MAKYPLEPVLAKKDRVDRAEKVYKEXRRLLEIQRKLEKEAE
RDKVKNHYMQIKQLDLDEGTTSDAVLQIKSYKVVAVQVSEEEKVKNQKEVYLA
ASKELEKAEVNLAKRRKEEKRHLHKEEMKEALKEARAEKEQDMLFQURQK
KKRESGGS"
8240..9082
/gene="CP0041"
8240..9082
/gene="CP0041"
/translation="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37936.1"
/db_xref="GI:7188981"
/translation="MELKKTAEESLYSAKTDNHTVYQNSPEPRDSRVKVFSLGKQTR
QERTTSKGNTRTESRFADEKRVDDIEAEVSGSEKQEQFCLAEAFAGMLSID
IAAAGSAAEVVEVAPVAVSDITQWNIENIILSTVESWVISEINGQLVELVDASSV
PEAFVGANLTLVQSGDLVSKFSSFVDATQMAEADLTNNPSQLSSLVLSALKGHQLT

```

Query Match

77.0%; Score 1098; DB 1; Length 25150;



Best local similarity 100.0%; Pred. No. 5, 6e-294;  
Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 ttgctgttaggaacccctctgacccaagcttataatgtgtgtacaaatggaaggt 188  
|||||  
Db 21138 TTGCTGTAGGAACCCCTCTGATCCAAAGCTTATTAATGTATGTTGATGGAAGGT 21069  
QY 189 gctgagagatccttgagatccttgagccttgctgacttggtgagcgtattagcttcgctgct 248  
|||||  
Db 21068 GCTGAGGAGATCCTTGGGATCCCTTGGCGCTACTTGGTGGCGAGCTATTAGCTTACGTGCT 21009  
QY 249 gattttacgagactatgtttgacacgtatcttataaagttagatgcacctaaacattt 308  
|||||  
Db 21008 GATTTTACGAGACTATGTTTTCACCGTATCTTAAAGTAGATGCACCTAAACAATT 20949  
QY 309 tctatggagccaagcctactggatccgctgctgctgcaactactactgacgtagataga 368  
Db 20948 TCTATGGGAGCAAGCTACTGGATCCGCTGCTGCAAACTACTACTGCGGTAGATAGA 20889  
QY 369 cctaaccgacctacaataagcattacacgatgcagagtggttcaactaaatgcaggttc 428  
|||||  
Db 20888 CCTAACCCGGCTACAATAAGCAATTTACGATGCGAGTGGTTCACCTAATGCAGGCTTC 20829  
QY 429 attgcttaaacatttgggagcgttttgatgttttctgacttttagggagcttctaattgt 488  
Db 20828 ATTGCTTTAAACATTTGGGATCCCTTTGATGTTTCTGCTTACTTTAGGAGCTTCTAATGCT 20769  
QY 489 tacattaggaatacctacacgcttcaactcgttggtttatttcgaggttaaaagtact 548  
Db 20768 TACATTAGAGAACTCTACACGCTCAATCTCGTGGTGTATTTCGGAGTTAAAGGTACT 20709  
QY 549 actgtaaatgcaatgaactacacacgctttctttaaagtaacgaggtgttgaactttac 608  
Db 20708 ACTGTAATGCAATGAACACTACCAACGTTCTTTTAAGTAACGAGTGTGTTGAACCTTAC 20649  
QY 609 acagacacctcttctcttctgagcgttaggcgctgctgagccttatgggaatgcggttgt 668  
Db 20648 ACAGACACCTCTTCTTCTTGGAGCGTAGCGCTCTGAGGCTTATGGGAATCGCGTTGT 20589  
QY 669 gcaacttggagctgaattccaattacacagtcacacgctcaaaccttaaaagtgaagaactta 728  
Db 20588 GCAACTTGGAGCTGAATTCGAATGACACAGCTCCAAACCTTAAGTTGAAGCACTTAT 20529  
QY 729 gtatctgtaacgtatcgcaattctctgtaaacacacccaaaggtctataaagcggtgt 788  
Db 20528 GTGATCTGAACGTATCGCAATTCCTGTAACAAACCCCAAGGCTATAAAGCGTGTCT 20469  
QY 789 ttcccttggcaacagacgctggcgttagcaacagctactggaacaaagtctgcgaccatc 848  
Db 20468 TTCCCTTGGCAACAGACGCTGGCGTAGCAACAGCTACTGGAACAAAGTCTGGACCATC 20409  
QY 849 aattatcatgaatggcaagtaggagcctctctctctctacagactaaactcttagtgcca 908  
Db 20408 AATTATCATGAATGGCAAGTAGAGAGCTCTCTATCTACAGACTAACTCTTTAGTGCCA 20349  
QY 909 tacattggagtacaatgggtctgagcaacttttgatgctgtataacacccgactgctcag 968  
Db 20348 TACATTGGAGTACAATGGTCTCGAGCAACTTTTGTATGCTGATAACTCGCAATTGCTCAG 20289  
QY 969 ccaaaactactacagctgttttaacttaactgaactggaacccctcttactaggaat 1028  
Db 20288 CCAAACTACTACAGCTGTTTTAACTTAACGCAATGAACCTCTCTTTACTAGGAAT 20229  
QY 1029 gccacagcattgctactactgattgcttctcagactctcatgcaaatgttttctgctcag 1088  
Db 20228 GCCACACATGCTACTACTGATGCTGCTCAGACTTCTATGCAAAATGTTCTCTCTCAG 20169  
QY 1089 atcaacaagtttaaatagaaaagcttggaggttactgtgagagactacttagtgat 1148  
Db 20168 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTTACTGTAGAGCTACTTTAGTTGAT 20109  
QY 1149 gctataaatggctactactcagagagcgtctttaaataacagagagcgtcgcagta 1208  
|||||

Db 20108 GCTGATAAATGGTCACTTACTGAGAGAGCTGCTTTAATTAACGAGAGCTGCTCAGTA 20049  
QY 1209 tctgggtcagttcagattc 1226  
|||||  
Db 20048 TCIGGTCAGTTTCAATTC 20031  
RESULT 6  
AP002547 300550 bp DNA BCT 08-DEC-2000  
LOCUS Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,  
DEFINITION section 3/4.  
ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008  
VERSION AP002547.2 GI:10176693  
KEYWORDS  
SOURCE Chlamydomophila pneumoniae J138 (strain:J138) DNA.  
ORGANISM Chlamydomophila pneumoniae J138  
REFERENCE 1 (sites) Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,  
Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.  
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWL029 from USA  
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
MEDLINE 2030349  
REFERENCE 2 (bases 1 to 300550)  
AUTHORS Shirai,M.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.  
Mutsunori Shirai, Yamaguchi University School of Medicine,  
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi  
755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,  
Tel:81-836-22-2227, Fax:81-836-22-2415)  
COMMENT On Sep 15, 2000 this sequence version replaced gi:6172298  
gi:6172300 gi:6172396 gi:6172398 gi:8978889.  
AB033786-AB033787, AB033816-AB033817: Submitted (25-Oct-1999).  
FEATURES  
source  
1. 300550  
/organism="Chlamydomophila pneumoniae J138"  
/db\_xref="taxon:138677"  
/note="synonym:Chlamydia pneumoniae (strain J138)"  
complement(21..794)  
/gene="dapP"  
complement(21..794)  
/gene="dapP"  
/codon\_start=1  
/transl\_table=1  
/product="diaminopimelate epimerase"  
/protein\_id="BAA98725.1"  
/db\_xref="GI:8978890"  
/translation="MAFYSPSTISKYFIYSGAGNRFLIGETLPEVEDRFLCQTRVD  
GFLYLPKSSCADAQLIIFNSDSRPTWCGNLRCALIAHLASQKGSIDISVSDGLYS  
GYLFYSDRLVDMTLADWRASVHRLSPDPPLKRVVCITHTGVPHAVILPEISTLDL  
SILGFLRYHOTFTSPDGVNPNVFOILGHCOLRVLYRVERVEGETAACGTGALASALVV  
SNSYGWKESIQIHTWGGELMTVSNQRGVYLGVSITRDL"  
complement(763..1338)  
/gene="clpP\_1"  
complement(763..1338)  
/gene="clpP\_1"  
/codon\_start=1  
/transl\_table=1  
/product="CLP protease"  
/db\_xref="GI:8978891"  
/translation="MADGEVHKLRDIEKELLEARRVFFSEPTVEKSDAIKKLWYL  
ELKDPKPIVFVINSPPGSDVDAWQIKNLTPVTVTGTAASMGSVLSLCAAP  
GRRATPHSRIMIHOPISGIPITGQATDLDIHAHELKTKARIIDVYVEATNQPRDII  
EKAIDRMWMTANEAKDPGLLDGILFSENDL"  
complement(1358..2917)  
/gene="glyA"  
complement(1358..2917)  
/gene="glyA"







CHTMOMB	1170 bp	DNA	BCT	26-APR-1993
LOCUS	Chlamydia pneumoniae major outer membrane protein (MOMP) gene,			
DEFINITION	complete cds.			
ACCESSION	M64064 M34942 M64063			
VERSION	M64064.1 GI:144534			
KEYWORDS	major outer membrane protein.			
SOURCE	Chlamydia pneumoniae (strain IOL-207) DNA.			
ORGANISM	Chlamydia pneumoniae			
REFERENCE	1. (bases 1 to 1170)			
AUTHORS	Carter, M.W., Al-Mahdawi, S.A.H., Giles, I.G., Trehan, J.D.,			
TITLE	Nucleotide sequence and taxonomic value of the major outer membrane			
JOURNAL	protein gene of Chlamydia pneumoniae IOL-207.			
MEDLINE	J. Gen. Microbiol. 137, 465-475 (1991)			
FEATURES	9123/311			
source	Location/Qualifiers			
	1. .1170			
	/organism="Chlamydia pneumoniae"			
	/strain="IOL-207"			
	/db_xref="taxon:83558"			
sig_peptide	1. .69			
	/gene="MOMP"			
CDS	1. .1170			
	/gene="MOMP"			
	/codon_start=1			
	/transl_table=11			
	/product="major outer membrane protein"			
	/protein_id="AA23143.1"			
	/db_xref="GI:144535"			
	/translation="MKLLKSLAFAFSGVLSQALPVGNPSDPSLLIDGTIWEGA			
	AGDPCDPCATWCDALISLRGFDYEDRLKVDAPKTFSGAKPTGSAANATTTAVD			
	RNPAYNKHLDHAEWTFNAGFIALINWDRDVFCTLGASNGYIRGNSAFNLVGLRGV			
	KGTIVANELPNVSLNGVVELYDTFSFVSGARGALWECGATLCAEQYQASLPK			
	VEELNVCNVSQFVSNPKYKGVAFPLPTDAGVATGTSATNVHEQVQASLSY			
	KLNSLPYIGVOWSRATFDADNIRIAPKLPVAVLNLTANNPSLIGNATLSTDSFS			
	DFMIVSCQINKFKSRACGVTVGATLVDAKDKSLTAEARLINEAHVSGQFRF"			
gene	1. .1170			
	/gene="MOMP"			
mat_peptide	70. .1167			
	/gene="MOMP"			
	/product="major outer membrane protein"			
BASE COUNT	319 a 255 c 247 g 349 t			
ORIGIN				
Query Match	76.9%; Score 1096.4; DB 1; Length 1170;			
Best Local Similarity	99.9%; Pred. No. 1.1e-293;			
Matches 1097;	Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY 129	ttgccttaggggaacccctctctgacccagcttatttaattgattgacgagcgttgcacataatgggaaggt 188			
Db 70	TTGCCTGTAGGGAACCCCTCTCTGATCCAGCTTATTAAATGATGGTACAAATATGGGAGGT 129			
QY 189	gctgcagagagatccttgcagctcttgcctacttgcctgctgctgagcagctattagcttgcgt 248			
Db 130	GCTGCAGGAGATCCTTGCCTACTTGCCTACTTGGTGCAGCGCTATTAGCTTACGTGCT 189			
QY 249	ggattttacggagactatgttttcgaccgtatcttaaaagtagatgcacctaacaattt 308			
Db 190	GGATTTTACGGAGACTATGTTTTCGACCGCTATCTTAAAAAGTAGATGCACCTAAAAATTT 249			
QY 309	tctatggggccaagcctactgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgt 368			
Db 250	TCATATGGGAGCCAAAGCCTACTGGATCCGCTGCTGCAAACTACTACTCCCGTAGATAGA 309			
QY 369	cctaaccggcgtcacaataagcatttacacgaagcagagtgctgagagtggttactaatgcaggttc 428			
Db 310	CCTAACCGCGCTTACAAATTAAGCATTTACAGATGCAGAGTGGTGTCTACTAATGACGCTTC 369			
QY 429	attgccttaaacattgggagcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc 488			



Db	370	ATTGCCCTTAACATTTTCGGATTCGCTTTTGATGTGTTTCCTGCTACTTTTAGGAGCGCTCTTAATGGT	429
Qy	489	tacattagagaaactctacagcgttcaactctcgtttggtttatttcgagtgataaaggctact	548
Db	430	TACATTTAGAGAAACTCTACAGCGTTCAATCTCGTTGGTTTATTTCGCGAGTTAAAGGTACT	489
Qy	549	actgtaaatgcaaatgaactaccaaagctttctttaaagtaaacggagtggttgaactttac	608
Db	490	ACTGTAATGCAATGAACATACCAACGGTTCTTTAAGTAACGGAGTTGTGTGAACCTTTAC	549
Qy	609	acagacacctcttctcttgagcgtgagcgtcgtgagccttatagggaatgcggttgt	668
Db	550	ACAGACACCTCTTTCTTGGAGCGTAGCGCTCGTGGAGCCTTATGSGAATCGGTTGT	609
Qy	669	gcaacttggagcgtgaattccaatatgcacagtcaccaactccaaactaaaagtgtgaagaactta	728
Db	610	GCAACTTTGGAGCTGAATTCCAATATGCACAGTCCAAACCTTAAAGTTGAAGAACTTAAT	669
Qy	729	gtgatctgtaacgctatcgcaattctctgttaacaaaccccaaggcgtataaaggcgttgt	788
Db	670	GTGATCTGTAAGCTATCGCAATTCTCTGTAAACAAACCCAAAGGCTATAAAAGCGTTGCT	729
Qy	789	ttcccttgcacacagcgtcgtgagtaacagcactctggaacaaagtctgagaccatc	848
Db	730	TTCCCTTGGCAACAGACGCTGCGGTAGCAACAGCTACTTGAACAAAGTCTGCAGCATC	789
Qy	849	aattatcatgaatggcgaagtagagcctctctatctcacagactaaaactcttagtgcca	908
Db	790	AATTATCATGAATGGCAAGTAGAGCCTCTATCTTTACAGACTAAACTCTTTAGTGCCA	849
Qy	909	tacattggatgacaatggtctcagcaactttgatgctgataacatccgcatgtctcag	968
Db	850	TACATTGGAGTACAAATGGTCTTCGAGCAACTTTTGATGCTGATAACATCCGATTGCTCAG	909
Qy	969	ccaaactactcagcgtttttaaacttaactgcagtcgaaccccttcttactaggaaat	1028
Db	910	CCAAACTACTACAGCTGTTTAAACTTTAACTTCATGCAATGGAACCCCTCTTTACTAGGAAT	969
Qy	1029	gccacagcattgtctactactgattcgtcttcagacttcgatgcaaatgtttctcgtctcag	1088
Db	970	GCCACAGCATTGCTACTACTGATTCTGCTCAGACTTCATGCAAAATGTTTCTCCTGTCAG	1029
Qy	1089	atacaagtttaaatctagaaaagctgtggagttactgtaggagctactttagtgtgat	1148
Db	1030	ATCAACAAGTTTAAATCTAGAAAGCTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT	1089
Qy	1149	gctgataaatggtcacctactctgcagaagctcgtttaaattaacgagagagctgctcacgta	1208
Db	1090	GCTGATAAATGGTTCACCTTACTGCAGAAGCTCGTTTAAATTAACGAGAGAGCTGCTCACGTA	1149
Qy	1209	tctggtcagttcagattc 1226	
Db	1150	TCTGGTCAGTTCAGATTTC 1167	
RESULT 8			
CPMOMPX			
LOCUS	CPMOMPX	1502 bp	DNA
DEFINITION	Chlamydomphila pneumoniae gene for major outer membrane protein.		
ACCESSION	X72023		21-MAY-1999
VERSION	X72023.1	GI:313844	
KEYWORDS	major outer membrane protein.		
SOURCE	Chlamydomphila pneumoniae.		
ORGANISM	Chlamydomphila pneumoniae		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomphila.		
AUTHORS	1 (bases 1 to 1502)		
TITLE	Girjes,A.A., Carrick,F.N. and Lavin,M.F.		
	Remarkable sequence relatedness in the DNA encoding the major outer		
	membrane protein of Chlamydia psittaci (koala type 1) and Chlamydia		
JOURNAL	pneumoniae		
MEDLINE	Gene 138 (1-2), 139-142 (1994)		
REFERENCE	94171025		
	2 (bases 1 to 1502)		

**AUTHORS** Lavin, M.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (22-APR-1993) M. Lavin, Professor of Molecular Oncology,  
 Queensland Institute of Medical Res, The Bancroft Centre, 300  
 Herston Road, Brisbane QLD 4029, AUSTRALIA  
**REFERENCE** 3 (bases 1 to 1602)  
**AUTHORS** Lavin, M.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (12-JUN-1993) M. Lavin, Professor of Molecular Oncology,  
 Queensland Institute of Medical Res, The Bancroft Centre, 300  
 Herston Road, Brisbane QLD 4029, AUSTRALIA  
**REMARK** updates [2]  
**FEATURES** Location/Qualifiers  
**source** 1..1602  
 /organism="Chlamydomophila pneumoniae"  
 /db\_xref="taxon:83558"  
 /note="biovar: koala; submitted as Chlamydia psittaci"  
 116..117  
 /citation=[2]  
 /replaces="t"  
 239..243  
 250..1419  
 /codon\_start=1  
 /transl\_table=11  
 /product="major outer membrane protein"  
 /protein\_id="CAA50906.1"  
 /db\_xref="GI:313845"  
 /db\_xref="SPTREMBL:Q08085"  
 /translation="MKLLKLSALLSAAFAGSVGSLQALPVGNPSDPSLLIDGTIWEIGA  
 AGDPCPCATDCAI SLRAGFYGDYVDFRIKLDVDPKFTSGMKAPGTGSATANTVTA  
 RNPAYNKKHLHDAEWFTNAGFYAIINLWDRDFEFTLGASNGYIKGNSAFNLVGLGF  
 KGTSVAAENLPNVLSNGVVELYTDTSFWSVGARGALWECGATCAEAGFYQAQSKPK  
 VELNLICNVAQFSVKNPKGYKGVAPLPDAGVATATGTSATINHEWQVGCASLYS  
 RLNSLPYIGVQMSRATEDADNRIIAQPKLPTAVLNLTAWNPSLIGNTTT"LATSDSPS  
 LFMQIVSCQINKFKSRKACGVTVGATLVLDADKSLTAERLINERAHVSGQFRF"  
 1450..1476  
 terminator 445 a 334 c 316 g 507 t  
**BASE COUNT** 445 a 334 c 316 g 507 t  
**ORIGIN**  
 Query Match 74.88; Score 1066; DB 1; Length 1602;  
 Best Local Similarity 98.2%; Pred. No. 3.3e-285;  
 Matches 1078; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
 Qy 129 ttgccttagggaaaccccttctgaccagcttataatgtatggtgacaaatgggaaggt 188  
 Db 319 TTGCCTTAGGGAACCTTCTGATCCAAGCTTATTATTGATGTACAAATATGGAAGGT 378  
 Qy 189 gctcagagagatccttcgcatccttgcgctacttggctgcgcgcgtatttagcttcgct 248  
 Db 379 GCTCAGAGAGATCCTTCGGAFCCTTGGCGCTACTTGGTGCAGCGCTATTAGCTTACGTGCT 438  
 Qy 249 ggatttcaggagactatgttttcgcgcgtatccttaaaagttagatgcacctaaacattt 308  
 Db 439 GGATTTTACGGAGACTATGTTTTCACCGTATCTTAAAGTAGATGCACCTAAACATTT 498  
 Qy 309 tctatggagccagcctacttgatccgctgcgctgcgaactatactatgcgtagataga 368  
 Db 499 TCTATGGAGCCAAACCTACTTGGATCGCGTACTTGCAACTACTACTTCTGCGGTAGATAGA 558  
 Qy 369 cctaaccgcgcctacaataagcatttacagatgcagatggttcactaatgcaggcttc 428  
 Db 559 CCTAACCTCGCTACAAATAGACATTTACACGATGCAGATGGTTCACCTAATGCAGGCTTC 618  
 Qy 429 attgccttaaacatttgggatacgtttgatgtttcttctgacttttaggagcttctaaggt 488  
 Db 619 ATTGCCTTAAACATTTGGGATCGCTTTGATGTTTCTGTACTTATAGGAGCTTCTAATGGT 678  
 Qy 489 tacattagagaaactctacaagcttcaactcgtttggtttatttcgaggttaagggtact 548  
 Db 679 TACATTAAAGGAACTCTACAGCGCTCAATCTCGTGTGTTTTATTTCGGAGTTTAAAGGTACT 738  
 Qy 549 actataaatgcgaatgaactaccacaaagctttctttaaagtaacgaggttgttgaactttac 608



CDS	/note="putative" 1. .1170 /gene="MOMP" /note="precursor; putative" /codon_start=1 /transl_table=11 /product="major outer membrane protein" /protein_id="AA17397.1" /db_xref="GI:289841" /translation="MKLLKSLAALSAAFAGSVGSQALPVGPNPSRPSLLIDGTIWGSA AGDCPCATWCDASLRAFGYVDVDFIRLIDAPKTEISMGAKPFGSATANTVTLVAV RNPAYNKLHYDAENFTNAGFYALNIWDFRFDVCTLGASNGYKGNAAFNLVGLFG KGTSNANPLPNVLSNGVIELYDFTWMSYGARGALWEGGCATLGAEFQVQAQSKPK VELNIVYCNVDSFSLNPKPGYGVAPPLPTDAGVNTAAGTKSATINHEWQVQASLKY RNSLVPIYIGVOWSRATFDADNIRIAQPLPTAILNLTAWNPSSLGSAATAVSSDDQFS DPMQIVSCQINKEFKSKKACGVTVGATLVADKWSLTAEARLINERAHISGQFRF" 70. .1167 mat_peptide									
BASE COUNT	312 a	246 c	251 g	361 t						
ORIGIN										
Query Match	69.9%	Score	997.2	DB 1	Length	1170				
Best Local Similarity	94.3%	Pred. No.	4.1e-266							
Matches	1039	Conservative	0	Mismatches	63	Indels	0	Gaps	0	
Qy	129	ttgcctgtaggaaccctctgatccaagcttataatgatgtgtacaatatgggaaggt	188							
Db	70	TTGCCTGTAGGAACCCCTCTGATCCAAGCTTATTAATGTATGTGTACAAATATGGGAAGGT	129							
Qy	189	gctcacagagatcctgcgacccctgcgctactgtgcgcactgtgcgcgctatttagcttactgct	248							
Db	130	GCTCAGAGAGATCCTTGGCATCCCTTGCGCTACTTGGTCGAGCGCTATTAGCTTACCTGCT	189							
Qy	249	ggatttcaggagactatgttttcgacgctatccttaaaagtagatgcacctaataaacattt	308							
Db	190	GGATTTTACGGAGACTATGTTTTCGACGCTATCTTAAAGATAGATGCCCTTAAACACTTT	249							
Qy	309	tctatggagccaagctactggatcccgctgcgcactactactactgcgtagataga	368							
Db	250	TCTATGGAGCCAAACCTACAGGATCCCGTACTGGCACTACTACCGCGCTAGATAGA	309							
Qy	369	cctaaccggcctaacaagcatttacacgatgcagatgtagtggttaccataatgcaggtctc	428							
Db	310	CCTAATCTCGCTACAAATAAGCATTTATACGATGCAGAAATGGTTTCACTAATCGAGCTTC	369							
Qy	429	attgcctlaaacatttggatgcgtttgaatgttttctgaacttttaggagcttctaaggt	488							
Db	370	ATTGCCTTAAACATTGGGATCGCTTTGATGTTTCTGTACCTTGGAGCTTCTAATGCT	429							
Qy	489	tacattaggaacctctacacggttcaactcgttttatttcggttatttcggtttaaagglact	548							
Db	430	TACGTTAAGGAACCTCTCGCGGCTCAATCTCGTTGGTTTATTTCGGAGCTTAAAGTACT	489							
Qy	549	actgtaaatgcataatgcatacacaacglttcttaagtaagcgaagctttgaactttac	608							
Db	490	TCGTAAATGCAATGAATCACTACCAACGGTTCTTTTAAGTAATGGAGTTATCGAGCTCTAT	549							
Qy	609	acagacacctcttctcttggagcgttagcgctcgttggagccttatgggaatcggttgt	668							
Db	550	ACAGATACHACTTTTCCTTGGAGCGTAGGTGCTCGTGGAGCTTTATGGGAATCGGTTGT	609							
Qy	669	gcaactttggagctgaaattccaatatgcacagtcacaaacctaaagttgaagaactaat	728							
Db	610	GCACATTGGGAGCTGAATTCCAATATGCACGCTCAACCTAAAGTTGAAGAACCTTAAT	669							
Qy	729	gtcatctgaactatcgcaattctctgtaacacaaaccccaagggctataaagcgcttgct	788							
Db	670	GTGATCTGTAACTATCGCAATTTCTTCTTAATAAACCCCAAGGCTATAAAGGCGCTTGCT	729							
Qy	789	ttcccttgcacaacagacgctggcgtagcaacagctactggaaacaaagctctggaccatc	848							







```

RESULT 11
AF131230      670 bp      DNA      BCT      01-JUN-1999
LOCUS      Chlamydomphila pneumoniae mutant major outer membrane protein (MOMP)
DEFINITION      gene, partial cds.
ACCESSION      AF131230
VERSION      AF131230.1 GI:4928269
KEYWORDS      Chlamydomphila pneumoniae.
SOURCE      Chlamydomphila pneumoniae.
ORGANISM      Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomphila.
REFERENCE      1 (bases 1 to 670)
AUTHORS      Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.
TITLE      Presence of viable Chlamydia pneumoniae in fetal calf serum and in
epithelial-derived cell lines
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 670)
TITLE      Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.
JOURNAL      Direct Submission
SUBMITTED      (25-FEB-1999) Pathology, Vanderbilt University, C-3321
MEDICAL CENTER North, Nashville, TN 37232, USA
FEATURES
source
1. 670
/organism="Chlamydomphila pneumoniae"
/db_xref="taxon:83558"
/note="synonym: Chlamydia pneumoniae
isolated from commercial fetal calf serum"
<1..>670
/gene="MOMP"
<1..>670
/gene="MOMP"
/translation="VDRPNPAYNKHLDHAEWFTNAGFTALNIWDFVDFCTLGASNGY
IRNSTAEFLVGLFGVGTNNANLPSVLSNGVVELYDTFSWSVGAKALWECG
CATLGAEFQYAGSKPEELVNCVQSVNKGKGVAPPLPTDGDVATATGTS
ATINHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAOPKPLTAVLNLTAWN
SLIGN"
variation
63
/gene="MOMP"
/note="difference from wild type, silent mutation"
/replaces="a"
BASE COUNT      191 a 147 c 138 g 194 t
ORIGIN
Query Match      46.9%; Score 668.4; DB 1; Length 670;
Best Local Similarity 99.9%; Pred. No. 9e-175;
Matches 669; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 360 gtaataagactaacccgctcaataagcattacacagtcagagtggttcaactaat 419
Db 1 GTAGATAGACTAACCCGGCTACATACATATACATTTACACGATGCAGAGTGGTTCACTAAT 60
Qy 420 gcaggcttcattgccttaaacattgggagtcgtttgatgtttttctgtacttttagagct 479
Db 61 GCCGGCTTCATTGCCTTAACATTTGGATCGCTTTCATGCTTTCTGTACTTTAGGAGCT 120
Qy 480 tctaattggtaccatagagaaactctacagcgttcaactctctgttttatttcgaggtt 539
Db 121 TCTAATGGTTACATATAGAGAACTCTACAGCGGTTCATCTCGTTGGTTTATTTCGGAGTT 180
Qy 540 aaaggctactctgtaaatgaatgaactaccacacgctttcttttaagcaacgaggttgtt 599
Db 181 AAAGGTACTACTGTAAATGCAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 240

```

```

Qy 600 gaactttacacagacacctcttctcttgagcgtagcgctgtagagccttattgggaa 659
Db 241 GAACTTTACACAGACACCTCTTCTTCTTGAGCGTAGGCGCTCGTGAGCCTTATGGAA 300
Qy 660 tgcggttgcaactttgggagctgaattccaatatgcacagtcacacccaaacccaaagttaa 719
Db 301 TCGGTTGTGCAACTTTGGAGCTGAATTCGAATATGCACAGTCCAAAGCTAAAGTTGAA 360
Qy 720 gaacttaattgctgtaacgctatcgcaattctctgtaaacacacccaaagggctataaa 779
Db 361 GAACTTAATGTGATCTGTAACGCTATCGCAATCTCTGTAAACAAACCCCAAGGCTATAAA 420
Qy 780 ggcgttgcttcccttgcacacagcgtggcgtgtagcaacagctactggaacaaagtct 839
Db 421 GCGTTCCTTCCCTTGCCACAGAGCTGCGTAGCAACAGCTACTGGAACAAAGTCT 480
Qy 840 ggcaccatcaattatcatgaatggcaagtaggagcctctctattcttacagactaaactct 899
Db 481 CGCACCATCAATTATCATGAATGCAAGTAGGAGCCTCTCTATCTTACAGACTAAACTCT 540
Qy 900 ttatgcccatacttgagtagacaatggtctcgagcaacttttgatgctgataacatccgc 959
Db 541 TTATGCGCATACATTGGATGATGCTCGAGCAACTTTTGATGCTGATAACATCCGC 600
Qy 960 atgctcagccaaactacacagcgtgttttaaaccttaactgcatgagaccccttcttta 1019
Db 601 ATTGCTCAGCCAAACTACCTACAGCTGTTTTTAACTTAACATGATGCAACCCCTTCTTTA 660
Qy 1020 ctaggaaatg 1029
Db 661 CTAGGAAATG 670

```

```

RESULT 12
AF131229      670 bp      DNA      BCT      01-JUN-1999
LOCUS      Chlamydomphila pneumoniae mutant major outer membrane protein (MOMP)
DEFINITION      gene, partial cds.
ACCESSION      AF131229
VERSION      AF131229.1 GI:4928267
KEYWORDS      Chlamydomphila pneumoniae.
SOURCE      Chlamydomphila pneumoniae.
ORGANISM      Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomphila.
REFERENCE      1 (bases 1 to 670)
AUTHORS      Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.
TITLE      Presence of viable Chlamydia pneumoniae in fetal calf serum and in
epithelial-derived cell lines
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 670)
TITLE      Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.
JOURNAL      Direct Submission
SUBMITTED      (25-FEB-1999) Pathology, Vanderbilt University, C-3321
MEDICAL CENTER North, Nashville, TN 37232, USA
FEATURES
source
1. 670
/organism="Chlamydomphila pneumoniae"
/db_xref="taxon:83558"
/note="synonym: Chlamydia pneumoniae
isolated from cryptically-infected H292 cells"
<1..>670
/gene="MOMP"
<1..>670
/gene="MOMP"
/translation="VDRPNPAYNKHLDHAEWFTNAGFTALNIWDFVDFCTLGASNGY
IRNSTAEFLVGLFGVGTNNANLPSVLSNGVVELYDTFSWSVGAKALWECG
CATLGAEFQYAGSKPEELVNCVQSVNKGKGVAPPLPTDGDVATATGTS
ATINHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAOPKPLTAVLNLTAWN
SLIGN"
variation
63
/gene="MOMP"
/note="difference from wild type, silent mutation"
/replaces="a"
BASE COUNT      191 a 147 c 138 g 194 t
ORIGIN

```

```

Query Match      46.9%; Score 668.4; DB 1; Length 670;
Best Local Similarity 99.9%; Pred. No. 9e-175;
Matches 669; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 360 gtaataagactaacccgctcaataagcattacacagtcagagtggttcaactaat 419
Db 1 GTAGATAGACTAACCCGGCTACATACATATACATTTACACGATGCAGAGTGGTTCACTAAT 60
Qy 420 gcaggcttcattgccttaaacattgggagtcgtttgatgtttttctgtacttttagagct 479
Db 61 GCCGGCTTCATTGCCTTAACATTTGGATCGCTTTCATGCTTTCTGTACTTTAGGAGCT 120
Qy 480 tctaattggtaccatagagaaactctacagcgttcaactctctgttttatttcgaggtt 539
Db 121 TCTAATGGTTACATATAGAGAACTCTACAGCGGTTCATCTCGTTGGTTTATTTCGGAGTT 180
Qy 540 aaaggctactctgtaaatgaatgaactaccacacgctttcttttaagcaacgaggttgtt 599
Db 181 AAAGGTACTACTGTAAATGCAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 240

```







Db	58	TTGCGCTAGGGAAACCCAGCTGAACCAAGTTTATTAAATCGATGGCACTATGTGGGAAGGT	117
Qy	189	qctgcagagatccttgcgatccttgcgtactcttggctgacgctattagcttaccgtgct	248
Db	118	GCTTCAGAGATCCTTCGGATCCTTGGCTACTTGGTGTGACGCCATTTAGCATCCGGCCA	177
Qy	249	ggattttacggagactatgttttgcgcgtatctttaaagtatagatgcacctaataacatt-	307
Db	178	GGATACTACGGAGATTATGTTTCGATCGTGATATAAAAGTTGATGTGAATAAACATATC	237
Qy	308	--ttctatggagccaagcctactggatccgctgctgcaactatactactgcccgtagat	365
Db	238	AGCGGATATGGGTGCAGTCTTACAGGAGGCCAGCGCATTTACAAACTCTCTACAGAT	297
Qy	366	agacctaaccqcgctacaataaagcatttacacgatgcagagtgttcaactaatgcaggc	425
Db	298	AGACCCACATCGCTTATATGGCAACACTTTCAGACGCTGAGTGGTTTACGAATCCAGCT	357
Qy	426	ttcattgcttaaacatttgggagctgtttagtgttttcttactattagagcttcaat	485
Db	358	TTCTTCGCGATTAAATATCTGGGATCGCTTTCATATTTCTGCATTAGTGTCTCCAAAT	417
Qy	486	ggttacattagagaaacctctacacgcttcaatctcgttggtttatttcggaggttaaaagt	545
Db	418	GGGTACTTCAAGCTAGTCTCTGCTGCATTCACCTCGTTGGTTGATTGGTGTAAAGGA	477
Qy	546	actactgtaaatgcaaatgaaactaccaaacgctttcttaagtaacggagttgtgaactt	605
Db	478	ACCTCGTAGCAGCTGATCAACTCCAAACGCTAGGCATCACTCAAGGTATTGTTGAGTTT	537
Qy	606	tacacagacacctcttctctggagctgagcgctcgtggagccttatggaaatcggt	665
Db	538	TACACAGATACAAATCTCTTTGGAGGTAGTGTGCACGTGGTGTCTTATGGGAATGTGCT	597
Qy	666	tgtgcaactttggagctgaaattccaatatgcaacgtccaaacctaaagtgaagaactt	725
Db	598	TGTGCAACTTTAGGAGCTGAATTCAGATGCTCAATCTAATCTCTAAATTCGAATGCTG	657
Qy	726	aatgtgatctgaagtatcgcaattctctgtaaacacccaagggtctataa-----a	779
Db	658	AATGTAATCTCCAGCCCAACAATTTGTAGTTTCAACGCCCTAGAGGATACAAAGGAACA	717
Qy	780	ggcgttgcttcccttgcacacagacgctggctgacacagctactggaacaaagtct	839
Db	718	GGATCGAACTTTCTCTTAACAGCTGGTACAGATGGTGTACAGATCTAAATCT	777
Qy	840	gcaacatcaattatcatgaatggaagttaggagcctctctcttacagactaaactct	899
Db	778	GCAACACTCAAAATATCATGAATGGCAAGTGGTTTACGCTCTCTTACAGATTGAACATG	837
Qy	900	ttagtccatcacattggagtcaaatggtctcgagcaactttgtatgctgataacatccgc	959
Db	838	CTTTGTTCCTTACATTTGGCGTTAACTGGTCAAGAGCAACTTTTGATGCTGACTATCCGC	897
Qy	960	attgctcagccaaactacactacagctgtttttaaacttaactgcatacgtggaacctcttta	1019
Db	898	ATCGCTCAACCTAAATTAGCCGCTGCTGTTTGAACCTTGACCATGTGACCACTCAATCT	957
Qy	1020	ctaggaatgccacagcattgtctactactgattcgttctcagacttcatgcgaattggt	1079
Db	958	TTAGGGGAAGCTACAGCTTTTAGATGTAGCAACAATAATCTCGGACTCTTTACAAATCGCT	1017
Qy	1080	tcctgtcagatcaacaagtttaactctagaaaaagcttgtggagtactcttagagactact	1139
Db	1018	TCGATTTCAGATCAACAAATGAATCTCAGAAAAGCTTGTGGTGTAGCTGTGTGGTCAACG	1077
Qy	1140	ttaagttagctgataaagtctacttactgcagaagctcgtttaataacagagagagct	1199
Db	1078	TTAATTCACGCTGACAAATGTCTAATCACTGGTGGTGAACACGCTTAATCAATGAAGAGCT	1137
Qy	1200	gctcagctatctggtcagttcagattc	1226
Db	1138	GCTTCATGAATGCTCAATTCAGATTC	1164

```

RESULT 14
AF272945      1170 bp      DNA      BCT      25-JAN-2001
LOCUS        Chlamydomophila abortus major outer membrane protein precursor (Ompl)
DEFINITION   gene, complete cds.

ACCESSION    AF272945
VERSION      AF272945.1 GI:12483891
KEYWORDS     Chlamydomophila abortus.
SOURCE       Chlamydomophila abortus
ORGANISM     Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE    1 (bases 1 to 1170)
AUTHORS      Vretou,E., Psarrou,E., Kaisar,M., Vlisidou,I., Saiti-Montesanto,V.
              and Longbottom,D.
TITLE        Identification of protective epitopes by sequencing of the major
              outer membrane protein gene of a variant strain of Chlamydia
              psittaci serotype 1
              Infect. Immun. 69 (1), 607-612 (2001)

JOURNAL      20569239
MEDLINE      11119563
PUBMED       2 (bases 1 to 1170)
AUTHORS      Longbottom,D. and Vretou,E.
TITLE        Direct Submission
JOURNAL      Submitted (29-MAY-2000) Moredun Research Institute, Bush Loan,
              Penicuik EH26 0PZ, UK

FEATURES     Location/Qualifiers
              1..1170
               /organism="Chlamydomophila abortus"
               /isolate="LLG"
               /specific_host="Capra aegagrus hircus"
               /db_xref="taxon:83555"
               /country="Greece"
               1..66
               /gene="Ompl"
               1..1170
               /gene="Ompl"
               /note="MOMP; porin"
               /codon_start=1
               /transl_table=11
               /product="major outer membrane protein precursor"
               /protein_id="AAG53881.1"
               /db_xref="GI:12483892"
               /translation="MKLLKSLILFRATGSALSALQALPYCNPAEPSILLIDGTWEGCAS
               GPDCPCSTQDAISIRAGYIGDFVDFRLKVDIKNTITGNGAVPTGTAARADKYKTPD
               RPNLAYGKHLDQAEWFNTAFALNINIDREDIFCTLGASNGYFKASASALNVLGLIG
               KGSSVAADSPAQITGIVVEFYDTTFSVSGARGALWEGCGATLGAEPYQAQSPK
               TELNVYSSPQAFVYHPRGYKGFAPLPLTAGTDQATDKSATIKYHEWQVGLALS
               RLNLVYPIGVNWSRPFDAIRIAOPKLAVALNLTWNPTLLGEATLDTNNKFA
               DFLOIASIQINMKSRKACGVAVGATLIDADKNSITGEARLINERAHHMNAQPRF"
               1..1170
               /gene="Ompl"
               67..1167
               /gene="Ompl"
               /product="major outer membrane protein"

BASE COUNT   322 a 263 c 255 g 330 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
```



```

187 GGATACGAGGAGATATGTTTCGATCGTGTATTAAGAGTGTATGAATAAAACTATC 246
QY 309 tc---tatggagcaagcactactgactcgctgctgcaaacataactactactccgtagat 365
Db 247 ACCGGCATGGGTGCGAGTTCCTACAGGAACCGGACGAGCTGTATACAAAACCTCCTACGGAT 306
QY 366 agacctaaacccgctacataaagcattacacgctgacagagtggttcaactaatcgagc 425
Db 307 AGACCAACATCGTTACGGCAACACCTTACAGAGCGCGAATGGTTACCAATCGAGCT 366
QY 426 ttcatggccttaaacatttggagctgttgatgtgttttctgacttttaggagcttcta 485
Db 367 TTCTCGCATTAATATCTGGGATCGCTTGTATATTTCTGCACATTTAGCGCTTCTAAT 426
QY 486 ggttcattagaggaaactcacgcttcaactcgttggttttattcgaggttaaggt 545
Db 427 GGGTACTTCAAGCTAGTTCTCGGCAATTTCAACCTCGTGGTGTGATTTGGTGTAAAGGA 486
QY 546 actactgtaaatgcaaatgaactacacaaacggtttctttaagtaacggaggtttgtaact 605
Db 487 TCCTCGTAGCAGCTGATCAGCTTCCCAATGTAGGCATCCTCAAGGAATCGTTGAATTT 546
QY 606 tacacagacacactcttcttcttggagcgtgagcgtctgtggagccttatgggaatcggt 665
Db 547 TATACAGATACAACTATCTTGGAGTGTAGTGCACGCGAGCTTTATGGAGTGTGT 606
QY 666 tgtgcaacttggagctgaattccaatatgcacagtcacaaacccctaaagtgaagaact 725
Db 607 TGTGCAACTTTAGGACGAGCTTCCAAATGAGCTGCTCTTATCCTTAAATTTGAATGTG 666
QY 726 aatgtgactgttaacgtatcgcaattctctgttaacaaacccaaagggctataaagcggt 785
Db 667 AATGTAGTCTCAGGCCAGCAAAATTTGTGGTTCACAAAGCTTAGAGATACAAAGGGAACA 726
QY 786 gcttcccttgcacacagacgctgctgtagcaacagctactggaacaaagctcgagacc 845
Db 727 GAATTTCCCTTACCTCTACAGCTGGTACTGTATGATCAGGCAACTGACACTAAGTCGGCTACA 786
QY 846 atcaattatcatgaatggcaagtagagcctctctctcttcttctcagactaaactcttagtg 905
Db 787 ATTAATACACCAATGGCAAGTTGGTTTACGCTCTCTTATCGATTTGAACATGCTTGT 846
QY 906 ccatacattggagtaacaaatgctgcagcaactttttagctgtagtaacacccgactgt 965
Db 847 CCTTACATTTGGCGTAAACTGGTCACGAGCAACTTTTGTATGCTGACGCTATCGGCATCGCT 906
QY 966 cagcctaaactcctcacagctgttttaaaacttaactgcatgaaacccctcttctactaga 1025
Db 907 CAACCTAATTTAGCTGCTGCTGTATTAACCTTACCAATGACCAATGGAACCCCAACCCCTTTTAGGG 966
QY 1026 aatgcaacagcattgtctactactgattgcttctcagacttcatgcaaatgttctcgt 1085
Db 967 GAAGCTACAACTTTAGATTAACCAACAAATTCGCTGACTTCTTGCAAATTTGCTCGATT 1026
QY 1086 cagatcaacaagtttaacttagaagaactgttgagttactgttaggagctacttagtt 1145
Db 1027 CAGATCAACAAATGAAGTCTAGAAAAGCTTGTGGTGTAGCTGTGGTGGTCAACCGTTAATC 1086
QY 1146 gatgctgataaatggtctactactgtagagcgtcgttttaataacgagagagcgtctcac 1205
Db 1087 GACGCTGACAAATGGTCAATCCTGTTGAAGCAGCCTTAAATCAATGAAGAGCGGCTCAC 1146
QY 1206 gtagctggtcagttcagattc 1226
Db 1147 ATGAATGCTCAATTCAGATTC 1167

```

RESULT 15  
 CHOMPAAID 1261 bp DNA BCT 07-MAY-1999  
 LOCUS Chlamydomophila abortus strain B577 major outer membrane protein  
 DEFINITION precursor (ompA) gene, complete cds.

M73036  
 M73036.1 GI:144564  
 Chlamydomophila abortus.  
 Chlamydomophila abortus  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
 1 (bases 1 to 1261)  
 Kaitenboeck, B., Kousoulas, K.G. and Storz, J.  
 Structures of and allelic diversity and relationships among the  
 major outer membrane protein (ompA) genes of the four chlamydial  
 species  
 J. Bacteriol. 175, 487-502 (1993)  
 93123168  
 Location/Qualifiers  
 1..1261  
 /organism="Chlamydomophila abortus"  
 /strain="B577"  
 /specific\_host="Ovis aries"  
 /db\_xref="taxon:83555"  
 /note="submitted as Chlamydia psittaci"  
 80..1246  
 /gene="ompA"  
 80..1249  
 /codon\_start=1  
 /transl\_table=11  
 /product="major outer membrane protein precursor"  
 /protein\_id="AAD29102.1"  
 /db\_xref="GI:4757612"  
 /translation="MKKLKLSALLEFAATGSAALSQAALPVGNPAPESLLIDGTMWEGAS  
 GDPDPCSTWCDATISIRAGYDYDFRVLVKNTITGMGAVPTGAAANAKTPTD  
 RNTAYGKHLQDAEWFNTAAFLANIMDFDICTFLGASNGYFKASSAFNLVLIGV  
 KPSSTAADQLPNVIGTIGTIVFTDTTFSWSVGARGALWEGCATLGAEFQVQSNPK  
 IEMLNVSPPAQFVHKPRGKGTAFPLPLTAGTDOATDKSATIKYHEWVGLALSY  
 RLNLMPYISVNSPRTADADARTQAPKLAALVNLTTWNPITLAGEATLDTSNKFA  
 DFLOIASIQINKMKSRKACGVAVGATLIDADKWSITGEARLINERAAHNAQFRF"  
 359 a 275 c 277 g 350 t

Query Match 42.1%; Score 600.2; DB 1; Length 1261;  
 Best Local Similarity 72.2%; Pred. No. 8.6e-156;  
 Matches 795; Conservative 0; Mismatches 303; Indels 3; Gaps 1;  
 QY 129 ttgctgtagggaacccctctgacacgcttattatgagtgagtaacatgggaaggt 188  
 Db 146 TTGCTGTAGGGAACCCAGCTGAACCAAGTTTATTATCATGATGGCACTATGTGGGAAGGT 205  
 QY 189 gctgcagagatccttgcgactctgctgctgctgctgctgctgctgctgctgctgctgct 248  
 Db 206 GCTTCAGGTGATCTTGGCATCTTGTCTACTTGTGTGATGCTATCAGCATCCCGCA 265  
 QY 249 ggattttacggagacatgttttcgacgctatctttaaagtagatgcacataaacattt 308  
 Db 266 GGATACACGAGATATGTTTTCGATCTGTATTAAGAGTTCATGTAATAAACATATC 325  
 QY 309 tc---tatgggagcaagcactgactgactgactgactgactgactgactgactgactgactgact 365  
 Db 326 ACCGGCATGGGTGCGAGTTCCTACAGGAACCGGACGAGCTTAATACAAAACTCCTACGGAT 385  
 QY 366 agacctaaacccgctacataaagcattacacgctgacagagtggttcaactaatcgagc 425  
 Db 386 AGNCCCAACATCGCTTACGGCAACACCTTACAGAGCGCGAATGGTTTACCAATCGAGCT 445  
 QY 426 ttcatggccttaaacatttggagctgttgatgtgttttctgacttttaggagcttcta 485  
 Db 446 TTCTCGCATTAATATCTGGGATCGCTTGTATATTTCTGCACATTTAGCGCTTCTAAT 505  
 QY 486 ggttcattagaggaaacactctcacgcttcaactcgttggttttattcgaggttaaggt 545  
 Db 506 GGGTACTTCAAGCTAGTTCTCGGCAATTTCAACCTCGTGGTGTGATTTAGTGTGTAAGGA 565  
 QY 546 actactgtaaatgcaaatgaactacacaaacggtttctttaagtaacggaggtttgtaact 605



Fri Feb 8 08:44:50 2002

us-09-391-606-12.rge

Db 566 TCCTCCATAGCAGCTGATCAGCTTCCCAATGTAGGCATCACTCAAGGAATCGTTGAATTT 625  
Qy 606 tacacagacaccttcttctcttgagcgtagcgctctgtgagccttatgggaatgcggt 665  
Db 626 TATACAGATACAACATCTCTTTGAGTGTAGTGCACGGGAGCTTTATGGAGTGTGGT 685  
Qy 666 tgtgcaactttgggagctgaattccaatatgcacagtcaccaaacctaaagtgaagaactt 725  
Db 686 TGTGGACTTTAGGAGCAGAGTTCCAATACGCTCAGTCTAATCTTAAATTTAAATGTTG 745  
Qy 726 aatgtgattctgtaacgtatcgcaattctctgtaaaaaaaccccaagggtataaaggcggt 785  
Db 746 AATGTAGTCTCCAGCCAGCACAATTTGTGTTTACAGCCCTAGAGGATACAGGGGAACA 805  
Qy 786 gctttcccttgcacaagcgctgqcgtagcaacagctactggaacaaagtctgcgacc 845  
Db 806 GCATTTCCCTTTACCTCTAACAGCTGGTACTGATCAGGCAACTGACACTAAGTCGGCTACA 865  
Qy 846 atcaattatcatgaatggcaagtagagccctctcttatcttacagactaaactctttagtg 905  
Db 866 ATTAATACACGAATGGCAAGTTGGTTTAGCGCTCTCTTTATCGATTGAACATGCTTGT 925  
Qy 906 ccatacattggagtaacaatggctctcgagcaacttttgaatgctgataacatcgcaattgct 965  
Db 926 CCTTACATTAGCGTAAACTGGTACAGAGCAACTTTTGTATGCTGACGCTATCGGCATCGCT 985  
Qy 966 cagccaaaaactacacagctgttttaaaacttaactgcatggaacccctcttactagga 1025  
Db 986 CAACCTAAATTAGCTGCTGCTGTGTTAACTTGACACATGGAGACCCCAACCTTTTAGGA 1045  
Qy 1026 aatgcacagcattgtctactactgattcgcttctcagacttcacgaattgttccctgt 1085  
Db 1046 GAAGCTACAGCTTTAGATACTAGCAACAATTCGCTGACTTCTTGCAAAATGCTTCGATT 1105  
Qy 1086 cagatcaacaagtttaaatctagaaaagcttggagttactgtgaggtactgttaggagcttagtt 1145  
Db 1106 CAGATCAACAAAATGAAGCTAGAAAAGCTTGTGGTGTAGTGTGGTGAACGTTAATC 1165  
Qy 1146 gatgtgataaataatggctacttactgcagaagctcgtttaattaaacgagagagctgtcac 1205  
Db 1166 GAGGCTGACAAAATGGTCAATCACTGGTGAAGCAGCGCTTAATCAATGAAGAGCCGCTCAC 1225  
Qy 1206 gtatctgggtcagttcagattc 1226  
Db 1226 ATGAATGCTCAATTCAGATTC 1246

Search completed: February 7, 2002, 19:37:19  
Job time: 13753 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 20:08:36 ; Search time 685.35 Seconds  
(without alignments)  
1783.827 Million cell updates/sec

Title: US-09-391-606-12

Perfect score: 1426

Sequence: 1 tgcagcgtactctgtctgc.....tccttgacctggaaggtgc 1426

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	600.2	42.1	1261	20 AAX25047	Chlamydia psittaci
2	526.4	36.9	1660	20 AAX25048	Chlamydia psittaci
3	523.2	36.7	1209	20 AAX25046	Chlamydia psittaci
4	514.6	36.1	1578	21 AAA64764	C. pneumoniae sero
5	514.6	36.1	1578	22 AAH56267	Chlamydia trachoma
6	479.4	33.6	3133	19 AAV62447	Chlamydia trachoma
7	479.4	33.6	3133	21 AAZ92753	DNA encoding Chlam
8	477.8	33.5	3133	19 AAV40646	Chlamydia trachoma
9	474.6	33.3	3133	7 AAN60007	Sequence encoding
10	397.2	27.9	1038602	20 AAX201425	Complete genome se
11	352.2	24.7	744	20 AAX25045	Chlamydia psittaci

12	328.8	23.1	1047	21 AAA08124	C. trachomatis MOM
13	321.8	22.6	726	20 AAX25044	Chlamydia psittaci
14	287.4	20.2	1362	21 AAA08122	C. trachomatis MOM
15	287.4	20.2	1371	21 AAA08120	C. trachomatis MOM
16	287.4	20.2	1542	21 AAA08123	C. trachomatis MOM
17	254	17.8	1452	21 AAA08121	C. trachomatis MOM
18	205.2	14.4	831	21 AAA08125	C. trachomatis MOM
19	171.6	12.0	1909	21 AAA27110	Rhesus monkey mela
20	140	9.8	720	7 AAN60006	Sequence encoding
21	140	9.8	720	19 AAV62446	Vector lambda gt11
22	140	9.8	720	19 AAV40645	Chlamydia trachoma
23	140	9.8	720	21 AAZ92752	DNA encoding Chlam
24	128	9.0	7475	20 AAX02659	EP-892047 Seq ID 3
25	128	9.0	8192	20 AAX02660	EP-892047 Seq ID 3
26	125	8.8	7108	20 AAX02662	EP-892047 Seq ID 3
27	120.4	8.4	4912	20 AAX08423	Recombinant vector
28	120.4	8.4	4912	21 AAA71402	Plasmid VR-1012 DN
29	120.4	8.4	5108	22 AAC86258	Plasmid GHRH1-29WT
30	120.4	8.4	5108	22 AAC86259	Plasmid GHRH1-29WT
31	120.4	8.4	5111	22 AAC86262	CMV IE promoter eh
32	120.4	8.4	5185	22 AAC86254	PGHRH-4 construct.
33	120.4	8.4	5188	22 AAC86266	Plasmid GHRH1-44YW
34	120.4	8.4	5254	22 AAC86267	Plasmid pGHRH1-44W
35	120.4	8.4	5322	21 AAA63225	Cytokine-encoding
36	120.4	8.4	5480	20 AAX81562	Plasmid VR4151, CO
37	120.4	8.4	5480	20 AAX81562	Cytokine-encoding
38	120.4	8.4	7001	20 AAX89795	DNA of pVR 1012-GP
39	120.4	8.4	7003	20 AAX59390	Plasmid pVR 1012-G
40	120.4	8.4	7073	20 AAX89796	DNA of pVR 1012-GP
41	120.4	8.4	7073	20 AAX59391	Plasmid pVR 1012-G
42	120.4	8.4	7285	20 AAX89797	DNA of pVR 1012-GP
43	120.4	8.4	7285	20 AAX59392	Plasmid pVR 1012-G
44	118.8	8.3	7272	20 AAX59393	Plasmid pVR 1012-s
45	118	8.3	4928	17 AAT32656	Plasmid encoding h

#### ALIGNMENTS

RESULT 1  
AAX25047  
ID AAX25047 standard; DNA; 1261 BP.  
XX AAX25047;  
AC AAX25047;  
DT 05-JUL-1999 (first entry)  
XX Chlamydia psittaci major outer membrane protein DNA.  
DE Major outer membrane protein; MOMP; psittacosis; infection;  
KW Vaccine; genetic immunisation; ss.  
OS Chlamydia psittaci.  
FH Key Location/Qualifiers  
FT CDS 80..1249  
FT /\*tag= a  
XX  
XX WO9910005-A1.  
XX  
XX PD 04-MAR-1999.  
XX  
XX PF 28-AUG-1998; 98WO-US17943.  
XX  
XX PR 28-AUG-1997; 97US-0057147.  
XX  
XX PA (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.  
XX  
XX PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
XX WPI: 1999-254214/21.  
XX DR P-PSDB; AAW98188.  
XX



CC This DNA sequence codes for the major outer membrane protein (MOMP),  
CC of *Chlamydia psittaci* strain 6BC. Claimed MOMP  
CC polypeptides (see AAW98187) of *Chlamydia psittaci* strain 6BC. Claimed MOMP  
CC polypeptides (see AAW98183 and AAW98184) comprise regions VD3 and VD4  
CC of an MOMP, i.e. they lack regions VD1 and VD2. Claimed vaccine  
CC compositions include such MOMP polypeptides, optionally fused to a  
CC maltose binding protein. Also claimed are isolated nucleic acids  
CC encoding the polypeptide, a vector, and a method of preventing C.



CC psittaci infection by administering the vaccine containing the  
CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking  
CC regions VD1 and VD2 are useful for genetic vaccination. The  
CC vaccines are used to prevent *C. psittaci* infection, especially in  
CC birds.

Sequence 1660 BP; 492 A; 342 C; 338 G; 488 T; 0 other;

Query Match 36.98; Score 526.4; DB 20; Length 1660;  
Best Local Similarity 69.58; Pred. No. 2.7e-153;  
Matches 792; Conservative 0; Mismatches 306; Indels 42; Gaps 4;

Qy	129	ttgcctctagggaacccttctgtaccgaagcttataattgtatggtacaatatgggaaggt	188
Db	430	ttgcctctagggaaccagctgaaccgaagtattataatcgatggcactatgtgggaaggt	489
Qy	189	gctcaggagatccctgcgacccctgcgctacttgggtgcacgcattagcttactgct	248
Db	490	gctcaggagatccctgcgacccctgcgctacttgggtgcacgcattagcatccgcgca	549
Qy	249	ggattttacggagactatgttttcgaccgctatcttataaagttagatgcacctaaacattt	308
Db	550	ggatctacgcgagattatgttttcgactcggtattataaagtgtgatataaactttt	609
Qy	309	tctatgggaccaaagcctactggatccgctgcgtcgaactatactact-----	356
Db	610	agcgcatggctgcacccctcacgcaggctacaggttaacgaagataactaatacagcca	669
Qy	357	---gccctagatagacctaaccgctcacataagaattttcacgcgtgcagagtggctc	413
Db	670	gaagcaaatggcagaccgaacatcgcttacggaaggcatatgcaagatgcagagtggctt	729
Qy	414	actaatcaggcttcattgccttataaaacatttgggatcgctttgtatgttttctgacttta	473
Db	730	tcaatgcagccttcctagccttataaaacatttgggatcgcttcgcacatttctgcacctta	789
Qy	474	ggagcttctaattggtttacattagaggaaactctcacgcgttccaactcgttgggtttatc	533
Db	790	gggcatccaatgatgacttcaaagcaagttcggctgcattccaacttggttgggttaata	849
Qy	534	ggagttaaaagt-----actactgtaaatgcaaatgaactaccacaacgcttct	581
Db	850	gggttttcagctgcaagctcaactcacttcacgcgttcacgtatctccaatgcacttctcaacgtaggc	909
Qy	582	ttaagtaacggagttgtgaactttacacagacaccttctctcttggagcgttagggcgt	641
Db	910	attaccaaaaggcttg-ggaaatttatacagacacacatttcttggagcgttaggtgca	969
Qy	642	cgtggagccttatgggaatgcggttggcaacttggagcgtgaattccaatttgcacag	701
Db	970	cgtggagccttatgggaatgfggttggcaactttagagcgtgagttccaatacgtcaca	1029
Qy	702	tccaacctaaagttgaagaacttaattgtatgctgaacgctatgcgaattctctgtaaac	761
Db	1030	tctaactctaagatgaatagtctcaacgctcacttcaagccacagcacaatttgtgatcac	1089
Qy	762	aaacccaagggtcataaaggcgtt-----gctttcccttgcacaacagcgtggcgta	815
Db	1090	aaacccaagggtcataaaggagctagctcgaatttcttcaactataacggctggaaca	1149
Qy	816	gcaacgctactgaaacaaagctctgcgaccataactatcatgaatggcaagttagggacc	875
Db	1150	acagaagctacagacaccaaatacgactacaaattataacacataatggcaagttaggcctc	1209
Qy	876	tctctaactcacagactaaactcttttagtgcatacatttggatgacaaatggtctcgagca	935
Db	1210	gccctgtctacaagattgaataatgctgtgtccataataattggcgttaactggcgaagagca	1269
Qy	936	actttgatgctgataacatccgcattgctcagccaaactacatcacagctgtttttaaac	995
Db	1270	actttgatgctgataactccgcatttgcctcaaccttaaatataaatacggagaattcttaac	1329

Qy	996	ttaactgcattgaacacctctcttactagaaatgccacagcattgtctactact-----	1049
Db	1330	attactacatggaacccaagccttataggatcaaccactgtttgccaataaatagtgtt	1389
Qy	1050	---gattcgtctcagaactctaatgcataattgtttctctgtcagatacaacaagttaaactt	1106
Db	1390	zaggatgtctctatctgatctctgcgaattgtctcgaattcagatcaacaagaatgaagtct	1449
Qy	1107	agaaaagccttgtaggttactctaggagctactcttagttgatctgataaaatggttcactt	1166
Db	1450	agaaaagccttgtaggttagctgttggtgcacagcttaatcgcagcgtgcacaaatggtcgaatc	1509
Qy	1167	actgcagaagcctcgttttaataacagagagagctgctcacgtatctcgttcagttcagattc	1226
Db	1510	actggttgagcagcgtttaatcaatgaagagctgctcagatgaatggtcacttcagattc	1569

RESULT 3  
AAX25046  
ID AAX25046 standard; DNA: 1209 BP.

AA  
AC

DT 05-JUL-1999 (first entry)

Chlamydia psittaci major outer membrane protein DNA.

KW Major outer membrane protein; MOMP; psittacosis; infection;

XX  
SECRET, SECURITY INFORMATION, NO FORN DISSEM

XX  
XX  
Columbus Period.

[illegible]

XX  
XX

XX  
XX

XX  
XX

XX  
XX[illegible]

DR P-PSDB; AAW98187.

PT A new vaccine for Chlamydia psittaci infections

PS Disclosure; Page 53-55; 72pp; English.

This DNA sequence codes for the major outer membrane protein (MOMP, see AAWS9187) of *Chlamydia psittaci* strain LSUW7CK, a cockatiel isolate (the MOMP gene sequence of this isolate is identical to that of *C. psittaci* Avian type C). A claimed MOMP polypeptide (see AAWS9188) comprises regions VD3 and VD4 of MOMP, i.e. it lacks regions VD1 and VD2. A claimed vaccine composition includes MOMP polypeptide lacking VD1 and VD2, optionally fused to a maltose binding protein. Also claimed are an isolated nucleic acid encoding the polypeptide, a vector, and a method of preventing *C. psittaci* infection by administering the vaccine containing the MOMP polypeptide. Vectors encoding MOMP polypeptides lacking regions VD1 and VD2 are useful for genetic vaccination. The vaccines are used to prevent *C. psittaci* infection, especially in birds.

Sequence 1209 BP; 352 A; 263 C; 260 G; 334 T; 0 other;

Query Match	Score	DB 20;	Length
Best Local Similarity	36.7%	523.2;	1209;
Matches 790; Conservative	69.3%	Pred. No. 2.2e-152;	
Matches 790; Conservative	0;	Mismatches 308;	Indels 42; Gaps 4;



QY	129	ttgccttaggaacccctctgatccaaagcttattaattgattggtacaatacggaaggt	188
Db	67		126
QY	189	gctcaggagatccttgcgatccttgcgctacttggtagcgagcctattagcttagctgct	248
Db	127		186
QY	249	ggattttacggagactatgttttcgacgctatcttaaaagttagtgcacctaaacattt	308
Db	187		246
QY	309	tctatgggagccaagcctactggatcgcgctgcgctgcgcaaacatactact	356
Db	247		306
QY	357	---gcgctagataacctaaccgctacaaataagcatttacacgatgcagagtgttc	413
Db	307		366
QY	414	actaatgcaggtcttattgcctttaaacaatttgggtagcgttttgattgttttcgtacttta	473
Db	367		426
QY	474	ggagcttctaattggtacattagaggaacctctacagcgttccaactcgtttggtttattc	533
Db	427		486
QY	534	ggagttaaaggt-----actactgtaaatgcgaatgaactacccaacggtttct	581
Db	487		546
QY	582	taagtacaagaggttgtgaacctttacacagacacctttctcttgagcgttaggcgt	641
Db	547		606
QY	642	cgtggagccttatgggaatcggttgtgcgaactttgggagctgaaattccaatatgcacag	701
Db	607		666
QY	702	tccaaacctaaagttagaagaacttaatgtgatctgtaacgtatgcgaattctctgtaaac	761
Db	667		726
QY	762	aaacccaagggctataaaagcgtt-----gotttcccttgcacaacgacgctggcgta	815
Db	727		786
QY	816	gcaacagctactggaacaaagtctgcgacatcaattatcatgaatggcaagttagagcc	875
Db	787		846
QY	876	tctctatttacagactaaactcttagtgccatacattgggagtacaa'ttggctotagaca	935
Db	847		906
QY	936	acttttgatcgtatacatccgcattgtctcagccaacaaactacctacagctgttttaaac	995
Db	907		966
QY	996	ttaactcatggaacctcttctactagggaatgcacagcatttgtctactact-----	1049
Db	967		1026
QY	1050	---gattcgtctcagactctaatgttttctcgttcagatcaacaagttaaatct	1106
Db	1027		1086
QY	1107	agaaaaagcttgtgagttactgtgagagctacttttagttgatcgttgataaattggtcaactt	1166
Db	1087		1146
QY	1167	actgcagaagctcgttttaattaaacgagagagctgcacgtctatctggtcagttcagattc	1226

Db	1147	actgttggaagcagcgttaataatgaagaagcgtctcacagaatgctcaattcagattc	1206
	RESULT	4	
	AAA64764		
	ID	AAA64764	standard; DNA; 1578 BP.
	XX		
	XX	AAA64764;	
	XX		
	DT	02-FEB-2001	(first entry)
	XX		
	DE	C. pneumoniae serovar MOMP5 pmp	gene Ral2 fusion coding sequence.
	XX		
	XX	Chlamydial infection; sexually transmitted disease;	
	KW	pelvic inflammatory disease; PID; tubal obstruction; infertility;	
	KW	trachoma; blindness; acute respiratory tract infection;	
	KW	heart disease; antibacterial; ss.	

### RESULTS

AAA64764 standard; DNA; 1578 BP.  
 AAA64764;  
 02-FEB-2001 (first entry)  
 C. pneumoniae serovar MOMPS pmp gene Ral2 fusion coding sequence.  
 Chlamydial infection; sexually transmitted disease;  
 pelvic inflammatory disease; PID; tubal obstruction; infertility;  
 trachoma; blindness; acute respiratory tract infection;  
 atherosclerosis; coronary heart disease; antibacterial; ss.  
 Chlamydia pneumoniae.  
 WO200034483-A2.  
 15-JUN-2000.  
 08-DEC-1999; 99WO-US29012.  
 08-DEC-1998; 98US-0208277.  
 08-APR-1999; 99US-0288594.  
 01-OCT-1999; 99US-0410568.  
 22-OCT-1999; 99US-0426571.  
 (CORI-) CORIXA CORP.  
 Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;  
 WPI; 2000-431303/37.  
 Isolated polypeptide for diagnosis and treatment of Chlamydia infection  
 comprises immunogenic portion of Chlamydia antigen, which comprises  
 amino acid sequence encoded by polynucleotide sequence -  
 Claim 1; Page 205; 256pp; English.  
 The present invention relates to new nucleic acid sequences and the  
 proteins encoded by the nucleic acid sequences. The encoded proteins  
 comprise an immunogenic portion of a Chlamydia antigen. The encoded  
 proteins are useful for the serodiagnosis and treatment of Chlamydia  
 infection. Chlamydiae are intracellular bacterial pathogens that are  
 responsible for a wide variety of human infections. C. trachomatis  
 infection is one of the most common sexually transmitted diseases and can  
 lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
 and infertility. Trachoma due to ocular infection with C. trachomatis is  
 the leading cause of preventable blindness worldwide. C. pneumoniae is a  
 major cause of acute respiratory tract infections in humans and is also  
 thought to play a role in the pathogenesis of atherosclerosis and  
 coronary heart disease. The present sequence is a nucleic acid sequence  
 isolated in the present invention.  
 Sequence 1578 BP; 414 A; 394 C; 382 G; 388 T; 0 other;

Query Match	36.1%	Score 514.6;	DB 21;	Length 1578;
Best Local Similarity	68.8%	Pred. NO. 1.2e-149;		
Matches 785; Conservative	0;	Mismatches 314;	Indels 42;	Gaps 4

Qy	128	gtgcctctgtagggaacccctctctgatccaaagcttatataattgatgtgacaaatatcggaag	247
Db	435	gtgcctctgtgggaatccagctgaacaaagcttatataatcgatggcaactatgtcggaag	494
Qy	188	tgcgcagaggagatcctctgcgatactctgcgatactttggtgcgacgcttatagcttacgtgc	247
Db	495	tgcctcaaggaaatcctctgcgatactctgcgatactttggtgcgacgcttatagcttacgtgc	554



RESULT	5
AAH56267	
ID	AAH56267 standard; DNA; 1578 BP.
XX	
AC	AAH56267;
DT	05-SEP-2001 (first entry)
XX	
Chlamydia trachomatis pmp gene sequence.	
XX	
Chlamydia; vaccine; infection; fusion protein; antigen;	
KW	pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
KW	acute respiratory tract infection; CapI; CT529; OMCB;
KW	polymorphic membrane protein; pmp; thiol specific antioxidant; TSA; ds.
XX	
Chlamydia trachomatis.	
OS	
XX	
WO200140474-A2.	
PN	
XX	
07-JUN-2001.	
PD	
XX	
04-DEC-2000; 2000WO-US32919.	
PF	
XX	
03-DEC-1999; 99US-0454684.	
PR	
19-APR-2000; 2000US-0556877.	
PR	
20-JUN-2000; 2000US-0598419.	
PP	
XX	
(CORI-) CORIXA CORP.	
PA	
XX	
Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;	
PI	
XX	
WIPI; 2001-374831/39.	
DR	
XX	
Chlamydia polypeptides and fusion proteins useful for preventing pelvic	
PT	inflammatory disease, trachoma, acute respiratory tract infections,
PT	atherosclerosis and heart disease -
P	
PS	Claim 1; Page 211; 295pp; English.
XX	
The present nucleotide sequence is provided in a specification	
CC	relating to compounds and methods for the treatment and diagnosis of
CC	chlamydial infection. The compounds provided include polypeptides and
CC	fusion proteins comprising immunogenic portions of Chlamydia antigens
CC	and DNA sequences encoding such polypeptides. They are useful for
CC	vaccinating against chlamydial infection, which causes pelvic
CC	inflammatory disease, trachoma, acute respiratory tract infections,
CC	atherosclerosis and heart disease.
XX	
Sequence 1578 BP; 414 A; 394 C; 382 G; 388 T; 0 other;	
SQ	
Query Match	36.1%; Score 514.6; DB 22; Length 1578;
Best Local Similarity	68.8%; Pred. No. 1.2e-149;
Matches	785; Conservative 0; Mismatches 314; Indels 42; Gaps
OY	128 gttgcctgtagggaaccccttctgatacagactatttaattgatggtacatatggagaag 187
Db	
435 gctgcctgtgggaatccaagctgaaccaagtattattaatcgatggcactatgtgggaagg 494	
OY	188 tgcctgcaggagatccttgcgatccttcgtcacttggtgcgacgctatttagcttacgtgc 247
Db	
495 tgcctcaggagatccttgcgatccttcgtcacttggtgcgacgctatttagcttacgtgc 554	
OY	248 tgcattttacgagactatgtttgacgcgtatcttaaagatagatgcacctaaacatt 307
Db	
555 agstactacgagatcatgtttcgatcgctgtataaagtgtgattgaataaacattt 614	
OY	308 ttctatgggagcaagcctactggatcccgctgctgcaaaactatactact----- 356
Db	
615 tagcgcatggtcgaaactcctcacgcgctataggtaacgcaagtaactaatcagcc 674	



Db 675 agaagcaaatggcagaccgaacatcgtcttaacggaagcatatgcaagatgcagatgggt 734  
 Qy 413 cactaagtgcagcttcattgcttaaacatttgatgctgcttcttctgctacttt 472  
 Db 735 ttcaaatgcagcttcattgcttaaacatttgatgctgcttcttctgctacttt 794  
 Qy 473 aggaagcttctaatgcttaccattagagaaactctacagcgttccaatctgctggtttatt 532  
 Db 795 agggatcccaatgatacttcaaacgaagctgcgttcattccaaactggttgggttaatt 854  
 Qy 533 cggagcttaaggt-----actactgtaaatgcaaatgaactacacaaacgtttc 580  
 Db 855 aggggtttcagctgcaagctcaatctctacagatcttccaaatgcaacttccaaacgtgag 914  
 Qy 581 tttaagtaacgaggtgtgtgaactttacacagacacactcttctctggagcgttaggcgc 640  
 Db 915 cattaccgaaggtgtgtggaattttatacagacacacatttcttggagcgttaggtgc 974  
 Qy 641 tctgtgagccttatgggaatgggtgtgtgcacacttggagcgttaattccaaatgacaca 700  
 Db 975 acgtggagctttatgggaatgggtgtgtgcacactttagagcgttaggtccaaatgacgtca 1034  
 Qy 701 gtccaaacttaagttgaagaacttaagtgtatctgttaacgtatcgcaattctctgttaa 760  
 Db 1035 atctaactcaagattgagatgtcaacgtcaacttcaagccacagcacaaatttggattca 1094  
 Qy 761 caaaccaaggctataaagggtt-----gctttcccttgcacacagcgttggcgt 814  
 Db 1095 caaaccaaggctataaagggtgctgctgaatttcttaccataaagcgttgggaac 1154  
 Qy 815 agcaacagctactgcaacaaagctgtgcacacacatttatcatgaatggcgaagtaggagc 874  
 Db 1155 aacagaagctacagacacacaaatcagctacaataataaccatgaatggcgaagtaggcct 1214  
 Qy 875 cctctatcttacagactaaactctttagtgcatacatttgagtcacaaatgtctcgcgc 934  
 Db 1215 cgcctctcttacagatgaatatctgttccatatttggcgttaaaactggtcgaagc 1274  
 Qy 935 aacttttgatgctataacatccgcattgctcagcacaactcactcagctggttttaa 994  
 Db 1275 aacttttgatgctatacctcgcattgctcacaactaaataaatacggagatctttaa 1334  
 Qy 995 cttaactgcatgaaccccttcttactaggaatgcaacagcattgtctactact- 1049  
 Db 1335 cattactatggaaccccaagccttataggatacaacacagctgttgcacaaataatggtg 1394  
 Qy 1050 ----gattcgttcagacttcacgaatttcttctgctcagatcaacaaatgtaaatc 1105  
 Db 1395 taaggatgtctatcgtatgcttgcacaaatgtcttcagatcaacaaatgaagtc 1454  
 Qy 1106 tagaaaagctgtgaggttactgtagagcactttagttagtgcgtgataaattgctact 1165  
 Db 1455 tagaaaagctgtggtgagctgtgtgcaacgtttaaactcagcgtgacaaatggtcatt 1514  
 Qy 1166 tactcagagcgtgtttaaataaacagagagcgtcgcacgtatctggtcagttcagatt 1225  
 Db 1515 cactggtgaagcagcgtttaaataaacgaagagcgtcgcacgtatggtcagaaatccgcgt 1574  
 Qy 1226 c 1226  
 Db 1575 c 1575

RESULT 6

AAV62447

ID AAV62447 standard; DNA: 3133 BP.

XX AC AAV62447;

XX AC AAV62447;

DT 02-FEB-1999 (first entry)

XX DE Chlamydia trachomatis major outer membrane protein DNA.

XX KW Major outer membrane protein; MOMP; diagnosis; vaccine; ds.  
 XX OS Chlamydia trachomatis serovar L2.  
 XX FH Key Location/Qualifiers  
 FT CDS 1288..2472  
 FT sig\_peptide /\*tag= a  
 FT mat\_peptide /\*tag= b  
 FT /\*tag= c  
 XX US5821055-A.  
 XX 13-OCT-1998.  
 XX 06-JUN-1995; 95US-0468451.  
 XX 13-JAN-1986; 86US-0818523.  
 XX 14-JAN-1985; 85US-0692001.  
 XX 25-APR-1991; 91US-0691839.  
 XX 28-OCT-1993; 93US-0144095.  
 XX 06-JUN-1995; 95US-0468451.  
 XX (CHIR ) CHIRON CORP.  
 XX (WASH-) WASHINGTON RES FOUND.  
 XX Agabian N, Kuo C, Mullenbach G, Stephens R;  
 XX WPI; 1998-567652/48.  
 XX P-PSDB; AAW73141.  
 XX Probe for detecting Chlamydia trachomatis - comprises polynucleotide  
 XX fragment that hybridises to major outer membrane protein DNA or RNA  
 XX Claim 5; Fig 2A-E; 15pp; English.

This DNA sequence codes for a 42 kDa major outer membrane protein  
 (MOMP, see AAW73141) of Chlamydia trachomatis serovar L2. A library  
 of chlamydial genomic DNA was produced in the phage lambda 1059  
 system. A lambda 1059 recombinant having a 9.2 kb insert was shown  
 to be homologous to lambda gt10/12/33 (see AAV62446) by Southern  
 analysis and was used for endonuclease digestion mapping and  
 additional Southern analyses. 2 Contiguous fragments were  
 identified and these contained sufficient base pairs to encode the  
 L2 MOMP gene product. These fragments were cloned into M13 for  
 DNA sequencing. Novel recombinant DNA constructs are provided for  
 the expression of a polypeptide having immunological activity  
 corresponding to that of a naturally-occurring MOMP of C.  
 trachomatis. Such polypeptides find use as reagents in the  
 detection of C. trachomatis, or antibodies to C. trachomatis, and  
 as vaccines against infection by C. trachomatis in susceptible  
 hosts. A claimed probe for detecting C. trachomatis comprises a  
 polynucleotide fragment that specifically hybridises to a DNA or RNA  
 sequence encoding C. trachomatis 38-45 kDa MOMP. The probe has  
 a sequence which is complementary to at least 12 contiguous bases  
 of the sequence given in AAV62447.

Sequence 3133 BP; 911 A; 667 C; 611 G; 944 T; 0 other;

Query Match 33.6%; Score 479.4; DB 19; Length 3133;  
 Best Local Similarity 66.5%; Pred. No. 1.6e-138;  
 Matches 741; Conservative 0; Mismatches 356; Indels 18; Gaps 3;

Qy 130 tgcctgtagggaacccttctcaatcgaagcttattgattggttgcacaaatgggaaggtg 189  
 Db 1355 tgcctgtagggaacccttctcaatcgaagcttattgattggttgcacaaatgggaaggtg 1414  
 Qy 190 ctgcaggagatccttgcgactccttgcctacttgcctgcttgcgagcgttattgcttgcgtg 249  
 Db 1415 tcggcggagatccttgcgactccttgcctacttgcctgcttgcgagcgttattgcttgcgtg 1474



```
QY 250 gatattacgagactatgttttcgaccgtatctctaaagtaagatgcacctaataacatttt 309
DB 1475 gtactatggtgactttgttttcgaccgtgttttgcacacagatgtgaataagaattcc 1534
QY 310 ctatggagcgaagccta-----ctggtccgctgctgcaaacataactactactgcgtag 363
DB 1535 aaatgggtgccaagcctaactgctacaggaatgctgcagctccatccactgtacag 1594
QY 364 atagacctaacccgctcacaataagcatttacagatgcagagtggttactaatatgag 423
DB 1595 caagagaaatcctgcttacgcccacatagcagatgctgagatgtttacaaatgtg 1654
QY 424 gcttcattgcttaaacatttgggtgctgtttgtgattgttttctgactttaggagcttcta 483
DB 1655 cttacatggcattgaatttgggtgctgtttgtgattgttttctgactttaggagcaca 1714
QY 484 atggttacattagagaaactctacagcgttcaatctggttgggttattcggagttta--- 540
DB 1715 gtggatatcttaaggaattcagatcttccaactgttggcttattcggagataatg 1774
QY 541 -----aaggtactactgtaaatgcaaatgaactaccacaaagctttcttaagtaacgag 594
DB 1775 agaaccatgctacagtttcagatagtaagctgttaccacaaatagagcttagatcaattg 1834
QY 595 ttgttgaactttacacagacacctttcttcttggagcgtgagcgtcgtgagccttat 654
DB 1835 ttgttgggtgtatcacagataacttttcttggagcgtgagcgtcgtgagccttat 1894
QY 655 ggaatgcggtgtgcaactttggagcgtgaattccaatagcagacagtcaccaaacctaaag 714
DB 1895 gggaaatggtgagcgtcgtttagcgtcttcttccaactacgctcaatcccaagcctaaag 1954
QY 715 ttgaagaacttaagtgtatgcttgaactgacgtatctcttgaacacaaacccaagcct 774
DB 1955 tgaagaattaaacttctgttaacgagcgtgagtttactatcaataaagcctaaagat 2014
QY 775 ataaaggcgttcttcccttgcacacagcgtggtggtgagcagactactggaacaa 834
DB 2015 atgtagggaagaattcccttcttgaacacagcgtggtggtgagcagacta 2074
QY 835 agtctgcgacacatcaattatcatgaatgagcagtaggagcctctcttctacagactaa 894
DB 2075 aggtatgctctattgattaccatgaatggaagcagtaggttagctctcttctacagacta 2134
QY 895 actctttagtgcatacatattgagtagcaatggtctcgagcaacttttgcgtgataaca 954
DB 2135 atatttcaactccctacattgagttaaatggtctcgagcaagtttgcgtgacacaga 2194
QY 955 tccgattgctcagcgaacaaactacacagcgtgttttaacttaactgcagtggaacctt 1014
DB 2195 ttcgtattgctcagcgaagtcagctacaactgtcttcttgcgttaccactctggaaccaa 2254
QY 1015 cttactagggaatgccacagcattgtctactactga---ttcgttctcagactctcagtc 1071
DB 2255 ctatgctgagcgtggcgtgtagaagctagcgcagagggctcagctcgagataccatgc 2314
QY 1072 aaattgttctcgtcagatcaacagtttaatacttagaagaagcgtgtgagttactgtag 1131
DB 2315 aaatggttctcgtcgaattgaacagatgaatcttagaagaatcttgcgttattgcagtag 2374
QY 1132 gagctactttagtgcataaattggttactactgcagaaagcgtctttaaataacg 1191
DB 2375 gaacaactattggtgagcagacaaatgcagtttgcagttgagactcgtctgacgtg 2434
QY 1192 agagagcgtcagctatctgtgacgttcagattc 1226
DB 2435 agagagcgtcagctaaatgcacattccgcttc 2469
```

RESULT 7  
AAZ92753  
ID AAZ92753 standard; DNA; 3133 BP.

```
XX AAZ92753;  
XX AC  
XX DT 19-JUN-2000 (first entry)  
XX DE DNA encoding Chlamydia trachomatis major outer membrane protein (MOMP).  
XX KW Major outer membrane protein; MOMP; serovar L2; immunoreactive;  
XX KW antibody production; immunoassay; detection; vaccine; trachoma;  
XX KW inclusion conjunctivitis; pneumonia; lymphogranuloma venereum;  
XX KW mucous membrane genital tract infections; ds.  
XX OS Chlamydia trachomatis serovar L2.  
XX FH  
XX FT Key Location/Qualifiers  
XX FT CDS 1288..2472  
XX FT /*tag= a  
XX FT /product= "Chlamydia trachomatis serovar L2 MOMP"  
XX PN US6030799-A.  
XX PD 29-FEB-2000.  
XX PF 06-JUN-1995; 95US-0466152.  
XX PR 28-OCT-1993; 93US-0144095.  
XX PR 13-JAN-1986; 86US-0818523.  
XX PR 25-APR-1991; 91US-0691639.  
XX PR 14-JAN-1985; 85US-0692001.  
XX PA (WASH-) WASHINGTON RES FOUND.  
XX PI Agabian N, Stephens R, Kuo C-C, Mullenbach G;  
XX DR WPI; 2000-223163/19.  
XX DR P-PSDB; AAY81266.  
XX PT New synthetic or recombinant polypeptide, useful for diagnosing or  
XX PT preventing Chlamydia trachomatis infection, is immunologically  
XX PT equivalent to a major outer membrane protein -  
XX PS Examples; Fig. 2; 17pp; English.  
XX CC The present sequence represents DNA encoding the major outer membrane  
XX CC protein (MOMP) of Chlamydia trachomatis serovar L2. DNA obtained from  
XX CC C. trachomatis serovar L2 was partially digested with Dnae I and  
XX CC inserted into the bacteriophage vector lambda-gt11; a lambda-gt11 insert  
XX CC (AAZ92752; ATCC #40157, referred to as lambda-gt11/L2/33 in the  
XX CC specification) was obtained which encodes a protein fragment (AAY81267)  
XX CC reactive with a pool of C. trachomatis-specific monoclonal antibodies.  
XX CC The lambda-gt11/L2/33 insert sequence was also used to probe a  
XX CC Chlamydia genomic DNA phage lambda 1059 library, and the present  
XX CC sequence which encodes the full-length Chlamydia trachomatis serovar L2  
XX CC MOMP was identified. C. trachomatis is a human pathogen responsible for  
XX CC diseases such as trachoma, inclusion conjunctivitis, pneumonia,  
XX CC lymphogranuloma venereum, and mucous membrane genital tract infections  
XX CC such as cervicitis and urethritis. The full-length recombinant MOMP or a  
XX CC fragment thereof is used to elicit the production of antibodies to a MOMP  
XX CC of Chlamydia trachomatis. MOMP proteins are also useful as immunoassay  
XX CC reagents (substitutes for native MOMP) for detecting C. trachomatis or  
XX CC its antibodies, for diagnosing infection, or as an immunogen for  
XX CC vaccines.  
XX SQ Sequence 3133 BP; 911 A; 670 C; 609 G; 943 T; 0 other;
```

Query Match 33.6%; Score 479.4; DB 21; Length 3133;  
Best Local Similarity 66.5%; Pred. No. 1.6e-138;  
Matches 741; Conservative 0; Mismatches 356; Indels 18; Gaps 3;

QY 130 tgcctgtagggaaccccttctgacccaagcttatttattggtgatacattatgggaaggtg 189  
DB 1355 tgcctgtgggaatcctctgacccaagccttattgacgacgaattctatgggaaggtt 1414



QY 190 ctgagagagactcttgatcctgctgctacttggtagcagcgtattagcttagctgctg 249  
 Db 1415 tcggcgagactcttgatcctgctgctacttggtagcagcgtattagcttagctgctg 1474  
 QY 250 gattttacgagagactatgttttcgacgctgctttaaagtagatgcacccaataatttt 309  
 Db 1475 gttactatggtgacttttttcgacgctggttttgcaaacagatgtgaataaagaattcc 1534  
 QY 310 ctatgggagcgaagccta-----ctggatccgctgtgtgcaaacctatactactcccgtag 363  
 Db 1535 aaatgggtgccaagcctaactgctacaggaatgctgcaagctccactccactgttagag 1594  
 QY 364 atagacctaacccgctacataaagcatttacacgagtagagtggttcaactaatcag 423  
 Db 1595 caagagagaatctgcttcaagcgcacatagcaggtgctgagatgtttacaatgctg 1654  
 QY 424 gtttcattgcttaaacatttggatcgtcttggatgttttctgacttttaggagcttcta 483  
 Db 1655 ctacatggcttgatgaatttgggagctgttttgatgtattctgtacattagagccacca 1714  
 QY 484 atggttacattagagaaactctacagcgttcaatctcgttgggtttatttcggagctta--- 540  
 Db 1715 gtgatattctaaaggaaattcagcatcttcaacttagttggcttatttcggagataatg 1774  
 QY 541 -----aaggtactactgataatgcaaatgaactccaacagttttcttaagtaacggag 594  
 Db 1775 agaaccatgctacagtttcagatagtaagctgttaccataatagcgttagatcaatctg 1834  
 QY 595 ttgtgaactttacacagacacactttctcttggagcgtgagcgtcgtgtagccttat 654  
 Db 1835 ttgttgagttgtatacagatactacttttcttggagtgctgagcgtcgtgtagcctttgt 1894  
 QY 655 gggaaatgcggttggtgcacatttgggagcgtgaattccaatatgcacagtcaccaacctaag 714  
 Db 1895 gggaaatggtgagcgcgactttagcgtcttccaatacgcctcaatccaagcctaag 1954  
 QY 715 ttgaagaacttaagtgtctgttaacgtatgcgaattctctgttaacaaaccccaaggct 774  
 Db 1955 tcgaagaattaaagttctctgttaacgcagcgtgagtttactataataagcctaaggat 2014  
 QY 775 ataaagcgtgtgttctcccttgcacacagcgtgctgctgtagcaaacagctactgaaaca 834  
 Db 2015 atgtagggcaagaattccctctgtatcttaagcaggaacagatggtgtgacaggaacta 2074  
 QY 835 agtcgcgacacataattatcatgaatggcaagtagagcctctctatcttcaacagactaa 894  
 Db 2075 aggatgctctattgattaccatgaatggcaagcagagtttagctctctcttcaacagactga 2134  
 QY 895 actctttagtgccatacattggagtaacaatggtctgcagcaactttttagctgtataaca 954  
 Db 2135 atattgtcaactcctacattggagtttaaatggtctgcagcaagtttttagtcagacacga 2194  
 QY 955 tcgcattgtcagcacaactcctacagcgtgttttaacttaacttaactgcatggaacccctt 1014  
 Db 2195 ttgattgtcagcgaagtcgctacaaactgtctttagtttaccactgtgaacccaa 2254  
 QY 1015 ctttactagaataatgcacagcattgtctactactga---ttcgttctcagacttctatgc 1071  
 Db 2255 ctattgtgagctggcagtgtaaaagctagcagagggctcagctgcgagataccatgc 2314  
 QY 1072 aaattgttctctgcagatcaacaagtttaaatctatgaaaaagcgttggagttactgctg 1131  
 Db 2315 aaatcgttctctgcgaattgaacaagatgaattctgaaaaatcttgcggtattgcagtag 2374  
 QY 1132 gagtactattagttgactgataaaatggtcacttactcagagcgtctgtttaaataag 1191  
 Db 2375 gaacaactattgtggtgcagacaaatcacgcagtcagtcagtcgttgcagtcgtgagatg 2434  
 QY 1192 agagagcgtcgcagctatcgttgctcagttcagattc 1226  
 Db 2435 agagagcgtcgcagctaaatgcacaaattccgcttc 2469

RESULT 8  
 AAV40646  
 ID AAV40646 standard; DNA; 3133 BP.  
 XX  
 AC AAV40646;  
 XX  
 DT 27-OCT-1998 (first entry)  
 XX  
 DE Chlamydia trachomatis serovar L2 MOMP coding region.  
 XX  
 KW MOMP; major outer membrane protein; immunoassay; diagnosis;  
 detection; antibody; serovar L2; ds.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 FH Key Location/Qualifiers  
 CDS 1288..2472  
 FT /\*tag= a  
 FT /product= MOMP  
 FT 1287..1353  
 FT sig\_peptide  
 FT /\*tag= b  
 XX  
 PN US5770714-A.  
 XX  
 PD 23-JUN-1998.  
 XX  
 PF 14-JAN-1985; 85US-0692001.  
 XX  
 PR 13-JAN-1986; 86US-0818523.  
 PR 14-JAN-1985; 85US-0692001.  
 PR 25-APR-1991; 91US-0691639.  
 PR 28-OCT-1993; 93US-0144095.  
 PR 06-JUN-1995; 95US-0466814.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (WASH-) WASHINGTON RES FOUND.  
 XX  
 PI Agabian N, Kuo C, Mullenbach G, Stephens R;  
 XX  
 DR WPI: 1998-376887/32.  
 DR P-PSDB; AAW57775.  
 XX  
 PT DNA coding for Chlamydia trachomatis polypeptide - useful for  
 producing recombinant polypeptide, etc.  
 XX  
 PS Claim 5; Fig 2; 15pp; English.  
 XX  
 CC The sequence is that encoding a major outer membrane  
 protein (MOMP) of Chlamydia trachomatis. This polypeptide can  
 be used in immunoassays, e.g. to detect Chlamydia trachomatis  
 antibodies in blood, or can be used in vaccines. The  
 polynucleotide can be labelled and used as a diagnostic probe.  
 XX  
 SQ Sequence 3133 BP; 909 A; 670 C; 610 G; 944 T; 0 other;

Query Match 33.5%; Score 477.8; DB 19; Length 3133;  
 Best Local Similarity 66.4%; Pred. No. 5e-138;  
 Matches 740; Conservative 0; Mismatches 357; Indels 18; Gaps 3;  
 QY 130 tcctctgtaggaacaccttctgacacgcttatttgaattggtgacacgcttattg 189  
 Db 1355 tcgctgtggtggggaatcctctgctgaacacgcttattgacgaggaattctatggaagtt 1414  
 QY 190 ctgcagggagatccttgcgactccttgcgactatttggtagcagcgtattgcttgcgtg 249  
 Db 1415 tcggcgagatccttgcgactccttgcacacgttggtagcgtatcagcatgcg-atgg 1474  
 QY 250 gattttacgagacactatgttttcgacgctattcttaaaagtagatgcacctaataatttt 309  
 Db 1475 gttactatggtgactttgttttcgacgctgttttgcacacagatgtgaataaagaattcc 1534



QY 310 ctatggggagccagccta-----ctggatccgctgtcgtcaaaactatactactgcccgtag 363  
 Db 1535 aaatgggtgccaagcctcaactgctacaggaatgctgcagctccatccactgtgacag 1594  
 QY 364 atagacctaacccgctcacaataaagcatttacaagatgcagtggttccactaatgcag 423  
 Db 1595 caagagagaatcctgcttaccgcccagacatgcagagctgctgagatgtttacaatagcgtg 1654  
 QY 424 gcttcattgcttaaaacttggagtcgcttggatgttttctgactctttagagagcttcta 483  
 Db 1655 ctacatggcattgaataattgggacgcttggatgttctgacatctcagagagccacca 1714  
 QY 484 atgggtacattagagaaactctacagcgttcaactcgttggatgttttattcgcgagttta--- 540  
 Db 1715 gggatattttaaagaaatcagcatcttccaactgttggcttattcgcgagataatg 1774  
 QY 541 -----aaggtactactgtaaatgcaaatgaactacccaacagtttctttaaagtaacggag 594  
 Db 1775 agaaccatgctacagtttcagatagtaagcttgcacaaatagatgagcttagatcaatctg 1834  
 QY 595 ttgttgaaactttacacagacacctcttctcttgagcgtgagcgtcgtgagccttat 654  
 Db 1835 ttgttgagttgtatacagatactacttcttgcttgagtgctgagagctcgtgagctttgt 1894  
 QY 655 gggaaatgggtgtgcacacttgggagctgaattccaatgcacagtcaccaacacctaaag 714  
 Db 1895 gggaaatgggtgcgagactttagcgcttcttccaactacgctcaatccaagcctaaag 1954  
 QY 715 ttgaagaacttaagtgtatgacttaacgtatgcgaacttctctgttaaaacaaacccaaggct 774  
 Db 1955 tgaagaataaaacgtctctctgtaacgcaggtgagtttactatcaataagagcctaaagat 2014  
 QY 775 ataaagcggtgtgttcttcccttgcacacagcgtgctgtagcaacagctacttgaacaa 834  
 Db 2015 atgtaggcgaagaattccctcttgatcttaagcaggaacagatggtgtgacaggaacta 2074  
 QY 835 agtctgcgaccatcaattatcaatgaatgcgaagtaggagcctctctctattctacagactaa 894  
 Db 2075 aggatgcctctattgtattaccatgaatggcaagcaagtttagctctctctctacagactga 2134  
 QY 895 actctttagtgcatacatgagtacaaatgctctgcagcaacttctgactgtatgactgaaca 954  
 Db 2135 atatgttactcctcatatggatgtaaatggtctcgcagcaagtttctgtagcagacaga 2194  
 QY 955 tcgcattgtcagcacaactacactacagctgtttttaaacttaactgcatggaacccctt 1014  
 Db 2195 ttcgtattgtcagcgaagtcagctacacactgtctttagttaccactctgaacccaa 2254  
 QY 1015 ctttactagaaatgccagcagatgtctactactga---ttcgttctcagacttcatgc 1071  
 Db 2255 ctattgtggagctggcgatgtgaaagctagcgcagaggggtcagctcggagataccatgc 2314  
 QY 1072 aaattgttctcctgcagatcaacaagtttaaatctagaagaagctgtgtgagttactgttag 1131  
 Db 2315 aaatcgttctcctgcgaatgcaacaagatgaatctctagaataatcttgcgttattgcagtag 2374  
 QY 1132 gagctacttagttgactgataaattgctactactactgcagaagcgttttaataaacg 1191  
 Db 2375 gaacaactattgtgagtcagacaaatgcagattacagttgagactcgtgactgactgagtag 2434  
 QY 1192 agagagctcactacgtatctgtgtcagttcagattc 1226  
 Db 2435 agagagctcactacgtataatgcacaaattccgcttc 2469

RESULT 9  
 AAN60007  
 ID AAN60007 standard; DNA; 3133 BP.  
 XX  
 AC AAN60007;  
 XX  
 DT 30-JUL-1991 (first entry)  
 XX

DE Sequence encoding a major outer membrane protein (MOMP) in  
 DE L2 B9-F DNA.  
 XX  
 KW Vaccines; trachoma; conjunctivitis; cervicitis; urethritis; ss.  
 XX Chlamydia trachomatis.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1288..1348  
 FT /\*tag= a  
 FT /product= leader  
 FT mat\_peptide 1349..2472  
 FT /\*tag= b  
 FT misc\_difference 2061..2063  
 FT /\*tag= c  
 FT /note= "AGA in AAN60006"  
 XX  
 PN EPI92033-A.  
 XX  
 PD 27-AUG-1986.  
 XX  
 PE 10-JAN-1986; 86EP-0100279.  
 XX  
 PR 14-JAN-1985; 85US-0692001.  
 XX  
 PA (CHIR-) CHIRON CORP.  
 XX  
 PI Agabian N, Stephens R, Kuo CC, Mullenbach GT;  
 XX  
 DR WPI; 1986-226702/35.  
 DR P-PSDB; AAP60004.  
 XX  
 DR New DNA constructs and polypeptide(s) - displaying antigenicity  
 PT of major outer membrane protein of Chlamydia trachomatis  
 XX  
 PS Example: Appendix B; 3lpp; English.  
 XX  
 CC DNA obt'd. from C. trachomatis serovar L2 was partially digested with  
 CC DNase I and inserted into vector lambda gt11. Clone lambda gt11/L2/  
 CC 33 (AAN60006) reacted with the pool and was subsequently shown to  
 CC produce a polypeptide that displays species-, subspecies- and type-  
 CC specific epitopes of the chlamydial MOMP. L2 B9-F DNA (AAN60007) is  
 CC comprised of lambda 1059 recombinants shown to be homologous with  
 CC lambda gt11/L2/33.  
 XX  
 SQ Sequence 3133 BP; 911 A; 673 C; 606 G; 943 T; 0 other;

Query Match 33.3%; Score 474.6; DB 7; Length 3133;  
 Best Local Similarity 66.2%; Pred. NO. 4.9e-137;  
 Matches 738; Conservative 0; Mismatches 359; Indels 18; Gaps 3;  
 QY 130 tgcctgtagggaacctctctgacacagcttatttgaatgattgataatggaaggtg 189  
 Db 1355 tgcctgtggggaatcctgctgaacaaagcttattgacgacgaattctatgggaaggtt 1414  
 QY 190 ctgcaggagatccttgcgactcttgcctactgttgcgacgcttattgacttgcgtg 249  
 Db 1415 tcggcgacatccttgcgactcttgcacactgttgcgactatcagatgcgtatgg 1474  
 QY 250 gattttaaggagactatgttttcgacctatctttaaagtagatgcacctaaacatttt 309  
 Db 1475 gttactatggtgacttgttttcgacctgttttgcacacagatgtgaataagaattcc 1534  
 QY 310 ctatgggagccagccta-----ctggatccgctgctgcaactatactactgcccgtag 363  
 Db 1535 aaatgggtgccaagcctcaactgctacaggaatgctgcagctccatccactgtgacag 1594  
 QY 364 atagacctaacccgctcacaataaagcatttacaagatgcagtggttccactaatgcag 423  
 Db 1595 caagagagaatcctgcttaccgcccagacatgcagagctgctgagatgtttacaatagcgtg 1654  
 QY 424 gcttcattgcttaaaacttggagtcgcttggatgttttctgactctttagagagcttcta 483



PN WO9528475-A2.  
XX  
PD 10-JUN-1999.  
XX  
PF 27-NOV-1998; 98WO-IB01939.  
XX  
PR 04-NOV-1998; 98US-0107077.  
PR 28-NOV-1997; 97FR-0015041.  
PR 17-DEC-1997; 97FR-0016034.  
XX  
PA (GEST ) GENSET.  
XX  
PI Griffais R;  
XX  
DR WPI; 1999-371125/31.  
PT Genome sequence of Chlamydia trachomatis  
XX  
PS Claim 1; Page 373-656; 1755pp; English.  
XX  
CC The present sequence represents the complete genome of Chlamydia trachomatis. Open reading frames (ORFs) of the genome encode polypeptides AY36754-Y37949. The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, perihepatitis, Bartholin's disease; lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.

Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 27.9%; Score 397.2; DB 20; Length 1038602;  
Best Local Similarity 66.4%; Pred. No. 1.3e-111;  
Matches 740; Conservative 0; Mismatches 357; Indels 18; Gaps 3

QY 130 tgctgtagggaacaccttctgatcccaagccttatattatggtgcataaatgggaagtg 189  
DB 733307 tgcctg9999aaactcctgtgaaccaagccttatgatcgacgagaattcttagggaagtt 733366  
QY 190 ctgaggagatccttcgatccctgcgtactctgtgcgacgctatagtctacgtctg 249  
DB 733367 tcggcggagatccttcgatccctgcacacctggtgtgacgctatcagcatcgctatgg 733426  
QY 250 gattttacggagactatgttttcgacgctatcttaaaagtagatgcacctaacaacatttt 309  
DB 733427 gtactatgtgactttgttttcgacgctatcttaaaagtagatgcacctaacaacatttt 733486  
QY 310 ctatgggagccaagccta-----ctggatcccgtcgtcgaactatactactcgcgtag 363  
DB 733487 aaatgggtgccaacctcacacactgctacaggaagcctgtcgagctccatcactgttacag 733546  
QY 364 atagacctaacccggcctacaataaagcattaccatgacagagtggtgttcactaatgtcag 423  
DB 733547 caagagagaatcctcgttcgacgacatgacagatgctcgagatgtttacaaaagcgtg 733606  
QY 424 gcttcattgccttaaacatttggtatcgttttctgtatgtttctgtactttaggagcttcta 483  
DB 733607 cttacatggcatgaaatttttggtatcgtttgtatgtattctgtactttaggagcaca 733666  
QY 484 atggttaccatagaggaactctcacgcgttcactcgtttcgttttatttcgaggtca--- 540  
DB 733667 gtgcatctcttaaaggaaattcagcatctctcaactctgtttggtttatttcgagagaaatg 733726  
QY 541 -----aaggtactactgtaaatgcgaatgaactaccacacggtttcttttaagttaacggag 594  
DB 733727 agaaccatgctacagtttcagatgaagcttgtaccacaaatagctgttagatcaatctg 733786  
QY 595 ttgttgactttacacagacacaccttcttcttcttctgtgagcgtgagcgctgctgagccttat 654

Db 1655 cttaacatggcattgaatttgggatcgttttgatgtattctgtacattaggaccacca 1714  
QY 484 atgggttacattagagaactctacagcgttcaactcgttgggtttatttcggagatta-- 540  
Db 1715 gtggatatcttaaggaataatcagcatcttcaacttagttggttatttcggagataatg 1774  
QY 541 -----aaggtactactgtaaatgcaaatgaactaccaaacggtttcttttaagtaacggag 594  
Db 1775 agaaccatgctacagtttcagatgtagctgttaccaaatatgagcttagatcaactcg 1834  
QY 595 ttgttgaaactttacacagacaccttcttcttcgttggagcgttagcgtcgtgagccttat 654  
Db 1835 ttgttgagttgatacagatactacttcttcttgggtgagtcgtgagcgtcgtgagccttctg 1894  
QY 655 gggaatcggtgtgcaacttttggagcgtgaattccaataatgcacagtccaacaccttaag 714  
Db 1895 gggaatgtagtcgcgagctttaggcgtctttccaataatgcacacacgctaag 1954  
QY 715 ttgaagaacttaatgtagctgtaacgtatgcgaattctctgttaaaaaaacccccaggtc 774  
Db 1955 tgaagaataaacttctctgtaacgcagctgagtttactatcaataaagcctaagat 2014  
QY 775 ataaagcgttcttcccttgccaacagacgcgtgcgtgacacagcactactggaacaa 834  
Db 2015 atgtagggcaagaattccctctgtattttaagcaggaacagatggtgtgacaggaacta 2074  
QY 835 agtcgcgaccatcaattatcatgaatggcaagttagggagcctctcttcttcaagactaa 894  
Db 2075 aggatcctctattgattaccatcaatggcaagcagtttagctctcttcaagactga 2134  
QY 895 acctttagtcatactatgagtaacatgctcgcagcaactttgtatgctgataaca 954  
Db 2135 atagtctactccactatgagttgaattgaatggctcgcgcaagttttgatgcagacaga 2194  
QY 955 tccgattgctcagcaaaactacctacagctggtttttaaacttaactgaactggaacctt 1014  
Db 2195 ttcgtattgctcagcgaagtcagctacacactcttctgtattgattaccactctgaaccaa 2254  
QY 1015 cttaactaggaatgccacagcattgtctactactga---ttcgttctcagacttcacgc 1071  
Db 2255 ctattgctgagctgcgagatgtgaagcgtgcgcgagaggtgcagctcggagataccacgc 2314  
QY 1072 aaattgttctcgtcagatcaacaagtttaaatctagaaaaagccttgcgttgcgttag 1131  
Db 2315 aaatcgttctcgtcgaatgaacagatgaatcagaaaaatctgcgttgcgttag 2374  
QY 1132 gagctacttttagttgctgataaatggtcacttactgacagaagcgtgtttaaataag 1191  
Db 2375 gaacaacttggatgcagacaaatgcgcagttacagttgagactgcgtgcgagc 2434  
QY 1192 agagagctctcagctatctgttcagttcagattc 1226  
Db 2435 agagagctctcagtaaatgcacaattcgcgttc 2469

RESULT 10  
ID AAZ01425  
AAZ01425 standard; DNA; 1038602 BP.  
XX  
AC AAZ01425;  
XX  
DT 07-OCT-1999 (first entry)  
XX  
DE Complete genome sequence of Chlamydia trachomatis.  
XX  
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
KW Bartholin's disease; lymphogranulomatosis; ss.  
XX  
OS Chlamydia trachomatis.

Accession	Sequence	Position
AAZ01425	atagacctaacccggcctacaataagcatttacagatgcagagtggttcactaatgcag	423
Qy		364
Db	caagagagaatcctgcttacggccgacatgcaggatgctgagatgtttacaaatgcg	733606
Qy		424
Db	cttacatggcattgaatatttggatcgtttgatgtattctgtacattaggagccacca	733666
Qy		484
Db	atggttacattagaggaactctacagctcaatcgtgtttatttcgagagta---	540
Qy		733667
Db	gtggtatcttaaaggaaattccagcatcttccaacttagttggttattcggagataatg	733726
Qy	-----aaggtactactgtaaatgcaaatgaactaccaaaacggtttcttttaagtaaacggag	594
Db	agaacctgctacagtttcagatagtaagctttgaccaaaatagcgtttagatcaaatctg	733786
Qy	ttgttgaaactttacacagacacctcttcttcttctggaagcgtgagcgctgctggagccattat	654







Db 616 tggtaactactgtgtaagcagcgttaataatgaagagccgctcactgaatgtctcaa 675

Qy 1218 ttcagattc 1226

Db 676 ttcagattc 684

RESULT 12

AAA08124  
ID AAA08124 standard; DNA; 1047 BP.

XX AC AAA08124;

XX DT 27-JUN-2000 (first entry)

XX C. trachomatis MOMP containing fusion protein nucleotide sequence #5.

XX Chlamydia trachomatis; fusion protein; major outer membrane protein;

KW MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;

KW infection; infectious disease; ds.

XX Chlamydia trachomatis.

OS Synthetic.

XX Key Location/Qualifiers

FT 1..1047

FT /tag=

FT /note= "Fusion protein containing at least part of a  
major outer membrane protein (MOMP) of Chlamydia  
trachomatis; no stop codon given"

XX JP2000041678-A.

XX 15-FEB-2000.

XX 28-JUL-1998; 98JP-0213212.

XX 28-JUL-1998; 98JP-0213212.

XX (ELED ) DENKI KAGAKU KOGYO KK.

XX WPI; 2000-295780/26.

XX P-PSDB; AAY82392.

XX A soluble fused protein useful for diagnosis of Chlamydia infection,

PT comprises at least part of major outer membrane protein (MOMP) of

PT Chlamydia trachomatis -

XX Claim 20; Page 29-30; 37pp; Japanese.

XX The present invention describes fusion proteins (I) comprising at least  
part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,  
at least one hydrophilic polypeptide having no immunoreactivity to  
human serum and their connected part. AAA08120 to AAA08125 encode  
specifically claimed examples of the fusion proteins given in AAY82388  
to AAY82393. Also described is a method (A) for the detection of  
Chlamydia trachomatis antibody using (I) as the antigen. (I) is used for  
the diagnosis of Chlamydia trachomatis infectious diseases. The method  
can diagnose Chlamydia trachomatis infectious diseases specifically in a  
high sensitivity.

XX Sequence 1047 BP; 315 A; 209 C; 233 G; 290 T; 0 other;

Query Match 23.1%; Score 328.8; DB 21; Length 1047;

Best Local Similarity 60.5%; Pred. No. 6.5e-92;

Matches 671; Conservative 0; Mismatches 367; Indels 72; Gaps 5;

Qy 126 atgttgctgtgaggaaacctctctgatcccaagcttattatgtatgttgataatgggaa 185

Db 1 atgtgctgtggttaacctctctgaaccaagcttattatgtatgtatgttgataatgggaa 60

Qy 186 ggtgctgagagatctctcgatccttgcgtactgtgcgtactgtgagacgtattactcgt 245

Db 61 ggttcggcggagatccttgatccttgccaccacttggtgagcgtatcagcatgcgt 120

Qy 246 gctggatttttaagagagactatgttttcgaccgtatcttaaaagtagatgcaactaaaca 305

Db 121 atgggttactatggtgactttgttttcgaccgtttttgcaaacagatgtaataaagaa 180

Qy 306 tttctatggagcgaagccta-----ctggatccgctgctgcaaacataactactgcc 359

Db 181 ttccaaatgggtgccaagcctcaactgctacaggcaatgctgcaagctccactcctgt 240

Qy 360 gtatagatagacctaacccgctcacaataagcatttacagcagtcagagtggttccaat 419

Db 241 acagcaagagagaatcctgcttaccgcccacatcagcagatgctgagatgtttacaa 300

Qy 420 gcaggcttcattgcttaaacatttgggacgttggatcgtttgattttctgacttttagagct 479

Db 301 gctgctacatggtcattgaatttgggacgtttgatttctgctacactcaggagcc 360

Qy 480 tctaaaggtttacattagaggaactctcagcgttccaactcgttttatttcggagtt 539

Db 361 accagtgatattcctaaggaatttcagcatcttccaacttagttggtttatttcggagat 420

Qy 540 aaaggttactactgtaaatgcaatgaactaccacaacgctttcttaagtaaacggagttgt 599

Db 421 aatg-----agaacctgctacagtttcagatagtagcgttgacca 462

Qy 600 gaactttacagacacactcttctcttgagcgtgagcgtcgttgagccttatggaa 659

Db 463 aatatgagcttagatcaatct----- 483

Qy 660 tgcggttgcaactttgggagctgaattccaatatgcagtcacagtcacaacccaagttgaa 719

Db 484 ----gtgttgagttgtatgttaacaataacaataacaataacaataacaataa 539

Qy 720 gaacttaattgctgatacgtatcgcaattctcttaaaacccaagcgtataaa 779

Db 540 caa--taacaataacaataacaatgctgagtttactatcaataagccttaagagataatgta 597

Qy 780 ggcgttgccttcccttgcacaagcgtggcgtagcaacagctacttgtaacaaagctct 839

Db 598 gggcaagaattccctcttgatcttaaaagcaggaacagatggtgacaggaactaaagat 657

Qy 840 ggcaccatcaattatcatgaatggcaagtaggagcctctctctacagactaaactct 899

Db 658 gcctctattgattaccatgaatggcaagcaggttagctctctctacagactgaatag 717

Qy 900 ttagtgccatacattggagtaacaatgltctcgagcaacttttgatgctgataacacatccgc 959

Db 718 ttcactccctacattggagttaaatggtctcgagcaagttttagtcagacagacttcgt 777

Qy 960 attgctcagccaaactacacagctgttttaacttaactgaactgcatggaacccctcttta 1019

Db 778 attgctcagccagagtcagctacacactgtctttagttaccactctgaacccaactatt 837

Qy 1020 ctaggaaatgccacagcattgtctactactga-----ttcgttctcagacttcacaaatt 1076

Db 838 gctggagctgcgctgattgaagcagcagagcagcagcgttcgagagataccatgcaaatc 897

Qy 1077 gttctctgcagatcaacaagtttaaatctagaaaaagcgttgaggttactgtaggagct 1136

Db 898 gttctctgcaattgaaacaagatgaaatctagaaaaatcttcggttattgcagtaggaaca 957

Qy 1137 actttagttgattgataaattggtcacttactgcagaagcgttttaatttaacagagaga 1196

Db 958 actattgtggatgcagacaaatcacagttacagttgagactcgtgtgatcgatgagaga 1017

Qy 1197 gctgctcagctatctggtcagttcagattc 1226

Db 1018 gctgctcagctaaatgcacaattccgcttc 1047



AA25044  
ID AAX25044 standard; DNA; 726 BP.  
AC AAX25044;  
XX  
DT 05-JUL-1999 (first entry)  
DE Chlamydia psittaci MOMP (minus VD1 and VD2 region) DNA.  
KW Major outer membrane protein; MOMP; psittacosis; infection;  
KW vaccine; genetic immunisation; ss.  
XX  
OS Chlamydia psittaci.  
FH Key Location/Qualifiers  
FT CDS 1..669  
FT /\*tag= a  
XX  
PN WO9910005-A1.  
XX  
PD 04-MAR-1999.  
XX  
PF 28-AUG-1998; 98WO-US17943.  
XX  
PR 28-AUG-1997; 97US-0057147.  
XX  
PA (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.  
PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;  
XX WPI; 1999-254214/21.  
DR P-PSDB; AAW98183.  
XX  
XX A new vaccine for Chlamydia psittaci infections  
XX  
XX Example 2; Page 43-45; 72pp; English.

This DNA sequence codes for a major outer membrane protein (MOMP) polypeptide (see AAW98183) of Chlamydia psittaci strain LSUWTKC, a cocktail isolate (the MOMP gene sequence of this isolate is identical to that of C. psittaci Avian type C). The MOMP polypeptide comprises regions VD3 and VD4 of native MOMP (see also AAW98187), i.e. it lacks regions VD1 and VD2 of MOMP. DNA encoding the MOMP polypeptide was obtained by PCR amplification (see also AAX25049 and AAX25051) of C. psittaci LSUWTKC DNA. A claimed vaccine composition includes this MOMP polypeptide, optionally fused to a maltose binding protein. Also claimed are an isolated nucleic acid encoding the polypeptide, a vector, and a method of preventing C. psittaci infection by administering the vaccine containing the MOMP polypeptide. Vectors encoding MOMP polypeptides lacking regions VD1 and VD2 are useful for genetic (naked nucleic acid) vaccination. The vaccines are used to prevent C. psittaci infection, especially in birds.

Sequence 726 BP; 221 A; 153 C; 144 G; 208 T; 0 other;

Query Match 22.6%; Score 321.8; DB 20; Length 726;  
Best Local Similarity 70.3%; Pred. No. 8e-90;  
Matches 466; Conservative 0; Mismatches 182; Indels 15; Gaps 2;  
QY 579 tctttaagtaacggaggtgttgaactttacacagacacacctcttctctggagcgtaggc 638  
DB 4 tccattaccacgaaggtgttgaagatttatacacacacacacattcttctggagcgtaggt 63  
QY 639 gctcgtggagccttgggaatgcggtgtgcaacttttgaggaggaattccaatgca 698  
DB 64 gcacgtggagccttgggaatgcggtgtgcaacttttgaggagcgtgtccaatgcgt 123  
QY 699 cagtcocaaacctaaagtgaagaacttaattgtatcgtatcgcgaattctctgta 758  
DB 124 caatctaactcctaagattgaatgtcgaacgtcacttccaagcgaagcaccattgtgatt 183

QY 759 acaaacaccaagggtctataaagcgtt-----gcttccctcttcccaacagacgctgac 812  
DB 184 cacaacaccaagggtctataaagcgtt-----gcttccctcttcccaacagcgtcga 243  
QY 813 gtacacacagctactggacaagaatctgcgaccataattatcatgaatgcaagtagga 872  
DB 244 acaacagaagctacagacaccaaatacagctacaataataaccataatggaagtaggc 303  
QY 873 gctctctatcttacagacataaactcttttagtgccatacattggagtagcaatggtctga 932  
DB 304 ctgcgcctgtcttacagattgaatatgttcttcataatttggcgttaactggtcaaga 363  
QY 933 gcaacttttgatgctgataacatccgcttgcgcagcccaaaactaccctacagctgttta 992  
DB 364 gcaacttttgatgctgatactatccgcttgcgcacacataataataatcgagattctt 423  
QY 993 aacttaactgcatggaaacctcttcttactaggaaatgccacagcatttgcctactact--- 1049  
DB 424 aacattactacatggaacccaagccttataggatcaacacactgcttgcaccaataatag 483  
QY 1050 -----gattcgtctcagacttcgaattgttctctgcagatacaacaagttaaa 1103  
DB 484 ggttaaggatgtctatctgatgtcttgcaaatgtctcagatcagatacaacaagaag 543  
QY 1104 tctagaaaagcttggagttactgtaggagctacttttagttgactgctgataaattggtca 1163  
DB 544 tctagaaaagcttgggtgtagctgttggtgcaacgttaatacgcgcgcacaaatggtca 603  
QY 1164 ctctactcagaagctcgttttaataacagagagcgtcgcacgtatctggtcagttcaga 1223  
DB 604 atcactggtgaagcagcgttcaatcaatgaagagcgtcgcacatgaatgctcaattcaga 663  
QY 1224 ttc 1226  
DB 664 ttc 666

RESULT 14  
AAA08122  
ID AAA08122 standard; DNA; 1362 BP.  
XX AAA08122;  
XX 27-JUN-2000 (first entry)  
DE C. trachomatis MOMP containing fusion protein nucleotide sequence #3.  
KW Chlamydia trachomatis; fusion protein; major outer membrane protein;  
KW MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;  
KW infection; infectious disease; ds.  
OS Chlamydia trachomatis.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT CDS 1..1362  
FT /\*tag= a  
FT /note= "Fusion protein containing at least part of a  
FT major outer membrane protein (MOMP) of Chlamydia  
FT trachomatis; no stop codon given".  
XX JP20000041678-A.  
XX 15-FEB-2000.  
XX 28-JUL-1998; 98JP-0213212.  
XX 28-JUL-1998; 98JP-0213212.  
XX (ELED ) DENKI KAKAKU KOGYO KK.  
XX WPI; 2000-295780/26.  
XX P-PSDB; AAY82390.







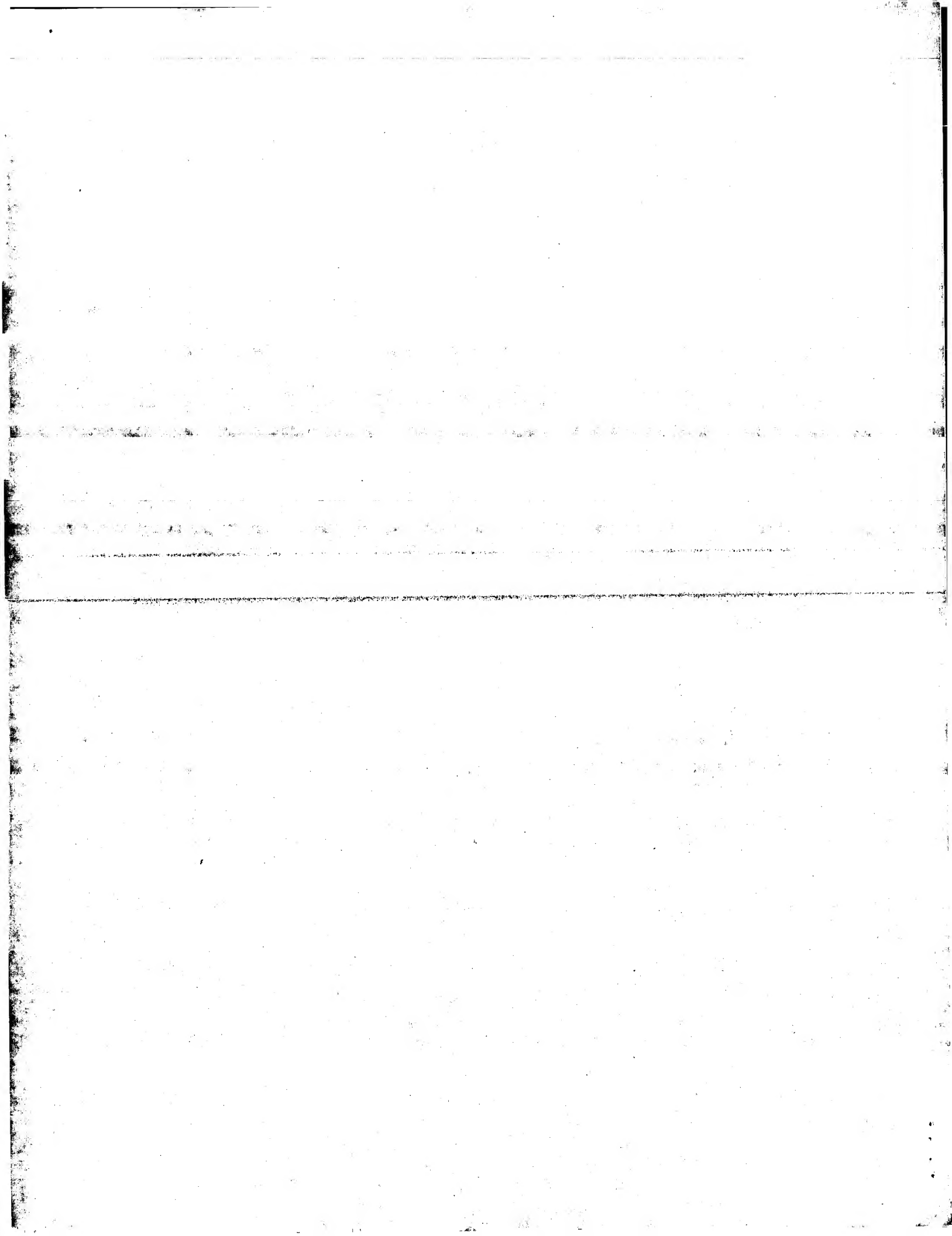
```

QY 829 gaacaaagtctggaccatcaattatcatgaatggcaagtagagcctctctatcttaca 888
Db 971 gaactaagagatgcctctattgattaccatgaatggcaagatttagctctcttaca 1030
QY 889 gactaaactcttttagtgccatcacattggagtagacaatggctcgagcaactttttagtgctg 948
Db 1031 gactgaatatgttcaactccctacattggagttaaaatggctcgagcaagtttttagtgca 1090
QY 949 ataacatccgcatctgctcagccaaactacacacagctgtttttaaaacttaactgcatgga 1008
Db 1091 acacgattcgtattgctcagccgaagtcagctacaactgtctttagttaccactctga 1150
QY 1009 accctcttactaggaatgccacagcattgtctactactga---ttcgttctcagact 1065
Db 1151 acccaactattgctggagctggcgatgtgaaagctagcgagagggcagctcggagata 1210
QY 1066 tcatgcaaatgtttctctgcagatcaacaagtttaaatctagaaaaagcttggaggtta 1125
Db 1211 ccattgcaaatcgtttctctgcaattgaacaagatgaaatctagaaaaatcttgcggtatg 1270
QY 1126 ctgtagagactctttagttgatgtgataaatggctcacttactgcagaagctcgttttaa 1185
Db 1271 cagtaggaacaactattgtggatgcagacaataacgcagttacagttgagactcgttga 1330
QY 1186 ttaacgagagagctgctcagctatctgttcagttcagattc 1226
Db 1331 tcgatgagagagctgctcagctaaatgcacaattccgcttc 1371

```

Search completed: February 7, 2002, 20:17:12  
Job time: 16146 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:54:07 ; Search time 243.49 Seconds  
(without alignments)  
1326.369 Million cell updates/sec

Title: US-09-391-606-12  
Perfect score: 1426  
Sequence: 1 tggacgactctgtgtgc.....tcctgacctggaagtgc 1426

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues  
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	8.3	4928	1 US-08-345-913-1	Sequence 1, Appl
2	118	8.3	4928	3 US-08-818-562-1	Sequence 1, Appl
3	104.8	7.3	4328	4 US-09-132-808-1	Sequence 1, Appl
4	104.8	7.3	4328	4 US-08-910-647-2	Sequence 2, Appl
5	104.8	7.3	4818	4 US-08-910-647-4	Sequence 4, Appl
6	104.8	7.3	5107	4 US-08-910-647-3	Sequence 3, Appl
7	104.8	7.3	9600	4 US-08-910-647-1	Sequence 1, Appl
8	104.2	7.3	3125	2 US-08-037-816A-13	Sequence 13, Appl
9	103.6	7.3	3125	2 US-08-530-146-13	Sequence 13, Appl
10	103.6	7.3	5676	2 US-08-663-998-3	Sequence 3, Appl
11	103.6	7.3	5682	2 US-08-663-998-4	Sequence 4, Appl
12	103.6	7.3	5900	2 US-08-663-998-1	Sequence 1, Appl
13	103.6	7.3	5952	2 US-08-663-998-2	Sequence 2, Appl
C 14	103	7.2	2057	1 US-08-450-945-57	Sequence 57, Appl
C 15	103	7.2	2057	4 US-08-976-161-57	Sequence 57, Appl
C 16	102.8	7.2	13254	1 US-08-276-852-156	Sequence 156, App
C 17	102.8	7.2	13254	1 US-08-276-852-170	Sequence 170, App
C 18	102.8	7.2	13254	1 US-08-899-575-156	Sequence 156, App
C 19	102.8	7.2	13254	1 US-08-899-575-170	Sequence 170, App
C 20	102.8	7.2	13254	1 US-08-899-575-156	Sequence 156, App
C 21	102.8	7.2	13254	1 US-08-899-575-170	Sequence 170, App
C 22	102.8	7.2	13254	5 PCT-US95-08743-156	Sequence 156, App
C 23	102.8	7.2	13254	5 PCT-US95-08743-170	Sequence 170, App
C 24	102.8	7.2	420	4 US-09-158-863C-64	Sequence 64, Appl
25	89.8	6.8	4326	4 US-08-760-615-7	Sequence 7, Appl
26	89.8	6.3	5253	2 US-08-627-151A-5	Sequence 5, Appl
27	86	6.0	3853	3 US-08-801-092-5	Sequence 5, Appl

28	86	6.0	4026	3 US-08-801-092-19	Sequence 19, Appl
29	86	6.0	4249	3 US-08-801-092-33	Sequence 33, Appl
30	86	6.0	4283	1 US-08-343-401A-3	Sequence 3, Appl
31	86	6.0	4283	1 US-08-445-265A-1	Sequence 1, Appl
32	86	6.0	4283	3 US-08-990-442-1	Sequence 1, Appl
33	86	6.0	4965	2 US-08-564-313-1	Sequence 1, Appl
34	86	6.0	4965	5 PCT-US94-06069-1	Sequence 1, Appl
35	86	6.0	5653	1 US-08-073-836-3	Sequence 3, Appl
36	86	6.0	5653	1 US-08-235-277-1	Sequence 1, Appl
37	86	6.0	6238	3 US-08-646-538-5	Sequence 5, Appl
38	86	6.0	6238	4 US-09-503-222-5	Sequence 5, Appl
39	86	6.0	6253	3 US-08-893-327-15	Sequence 15, Appl
40	86	6.0	6280	3 US-08-893-327-17	Sequence 17, Appl
41	86	6.0	6280	3 US-08-893-327-19	Sequence 19, Appl
42	86	6.0	6367	1 US-08-470-299-1	Sequence 1, Appl
43	86	6.0	6367	3 US-08-778-511-3	Sequence 3, Appl
44	86	6.0	6926	1 US-08-470-299-2	Sequence 2, Appl
45	86	6.0	7106	1 US-08-188-281B-8	Sequence 8, Appl

## ALIGNMENTS

RESULT 1  
US-08-345-913-1  
Sequence 1, Application US/08345913  
Patent No. 5641665  
GENERAL INFORMATION:  
APPLICANT: Hobart, Peter  
APPLICANT: Parker, Suzanne  
APPLICANT: Margalith, Michal  
APPLICANT: Khatibi, Shirin  
TITLE OF INVENTION: PLASMIDS SUITABLE FOR IL-2 EXPRESSION  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/345,913  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ways Vensko, Nancy  
REGISTRATION NUMBER: 36,298  
REFERENCE/DOCKET NUMBER: VICAL.043A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4928 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
FEATURE:



RESULT 3  
US-09-132-808-1  
; Sequence 1, Application US/09132808  
; Patent No. 6197332  
; GENERAL INFORMATION:  
; APPLICANT: Ronald Zuckermann et al.

```

RESULT#      4
US-08-910-647-2
; Sequence 2, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.3C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:

```



```

Query Match          7.3%; Score 104.8; DB 4; Length 4818;
Best Local Similarity 94.0%; Pred. No. 1.4e-23;
Matches 109; Conservative 0; Mismatches 7; Indels 0; Gaps
QY 1 tsgacagtactcgttctgcgcgcgcgcacacagacataatagctgcagactaacaga 60
      |||||
Db 1549 TGAGCAGTACTCGTCTGCTCGCGCGCGCCACACACATAATAGCTGCAGACTAACAGA 1608

QY 61 ctgttcctttccatgggtttttctgcagtcaccctgcctgcagacgctgtgatcaga 116
      |||||
Db 1609 CTGTTCCTTTCCATGGGTCTTTTCTGCAGTCACCGCTCGACCTAAGAATTCAGA 1664

RESULT 6
US-08-910-647-3
; Sequence 3, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-910-647-3

Query Match          7.3%; Score 104.8; DB 4; Length 5107;
Best Local Similarity 94.0%; Pred. No. 1.4e-23;
Matches 109; Conservative 0; Mismatches 7; Indels 0; Gaps
QY 1 tsgacagtactcgttctgcgcgcgcgcacacagacataatagctgcagactaacaga 60
      |||||
Db 1549 TGAGCAGTACTCGTCTGCTCGCGCGCGCCACACACATAATAGCTGCAGACTAACAGA 1608

QY 61 ctgttcctttccatgggtttttctgcagtcaccctgcctgcagacgctgtgatcaga 116
      |||||
Db 1609 CTGTTCCTTTCCATGGGTCTTTTCTGCAGTCACCGCTCGACCTAAGAATTCAGA 1664

RESULT 7
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:

```







Qy 1 tgägcagtgactcgttgctgccgcgcgcgcgcacacataatagctgacagactaacaga 60



Db 1584 TGAGCAGTACTCGTTGCTGCGCGCGCCACACAGACATAATAGCTGACAGACTAACAGA 1643  
Qy 61 ctgttcctttccatgggtcttttgcagtcacccgtgacacacgtgtgatcagatatc 120  
Db 1644 CTGTTCCCTTCATGGGCTCTTTTGTGAGTCACCGTCGTCGACACAGAGCTGAGATCCTAC 1703  
Qy 121 ccacatgttgccttagggaaacctctg 150  
Db 1704 AGGAGTCCAGGGCTGGAGAGAAACCTCTG 1733

RESULT 12  
US-08-663-998-1  
; Sequence 1, Application US/08663998  
; Patent No. 5846946  
; GENERAL INFORMATION:  
; APPLICANT: HUEBNER, Robert C.  
; APPLICANT: NORMAN, Jon A.  
; APPLICANT: LIANG, Xiaowu  
; APPLICANT: CARNER, Kristin R.  
; APPLICANT: BARBOUR, Alan G.  
; APPLICANT: LUKE, Catherine J.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/663,998  
; FILING DATE: 06-JUN-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kowalski, Thomas J.  
; REGISTRATION NUMBER: 32,147  
; REFERENCE/DOCKET NUMBER: 454312-2440  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-840-3333  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5900 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-663-998-1

Query Match 7.3%; Score 103.6; DB 2; Length 5900;  
Best Local Similarity 80.7%; Pred. No. 3.7e-23;  
Matches 121; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
Qy 1 tgagcagtactcgttgcgcgcgcgcacacacacataatagctgacagactaacaga 60  
Db 1588 TGAGCAGTACTCGTTGCTGCGCGCGCCACACAGACATAATAGCTGACAGACTAACAGA 1647  
Qy 61 ctgttcctttccatgggtcttttgcagtcacccgtgacacacgtgtgatcagatatc 120  
Db 1648 CTGTTCCCTTCATGGGCTCTTTTGTGAGTCACCGTCGTCGACACAGAGCTGAGATCCTAC 1707  
Qy 121 ccacatgttgccttagggaaacctctg 150  
Db 1708 AGGAGTCCAGGGCTGGAGAGAAACCTCTG 1737

RESULT 13  
US-08-663-998-2  
; Sequence 2, Application US/08663998  
; Patent No. 5846946  
; GENERAL INFORMATION:  
; APPLICANT: HUEBNER, Robert C.  
; APPLICANT: NORMAN, Jon A.  
; APPLICANT: LIANG, Xiaowu  
; APPLICANT: CARNER, Kristin R.  
; APPLICANT: BARBOUR, Alan G.  
; APPLICANT: LUKE, Catherine J.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/663,998  
; FILING DATE: 06-JUN-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kowalski, Thomas J.  
; REGISTRATION NUMBER: 32,147  
; REFERENCE/DOCKET NUMBER: 454312-2440  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-840-3333  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5952 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-663-998-2

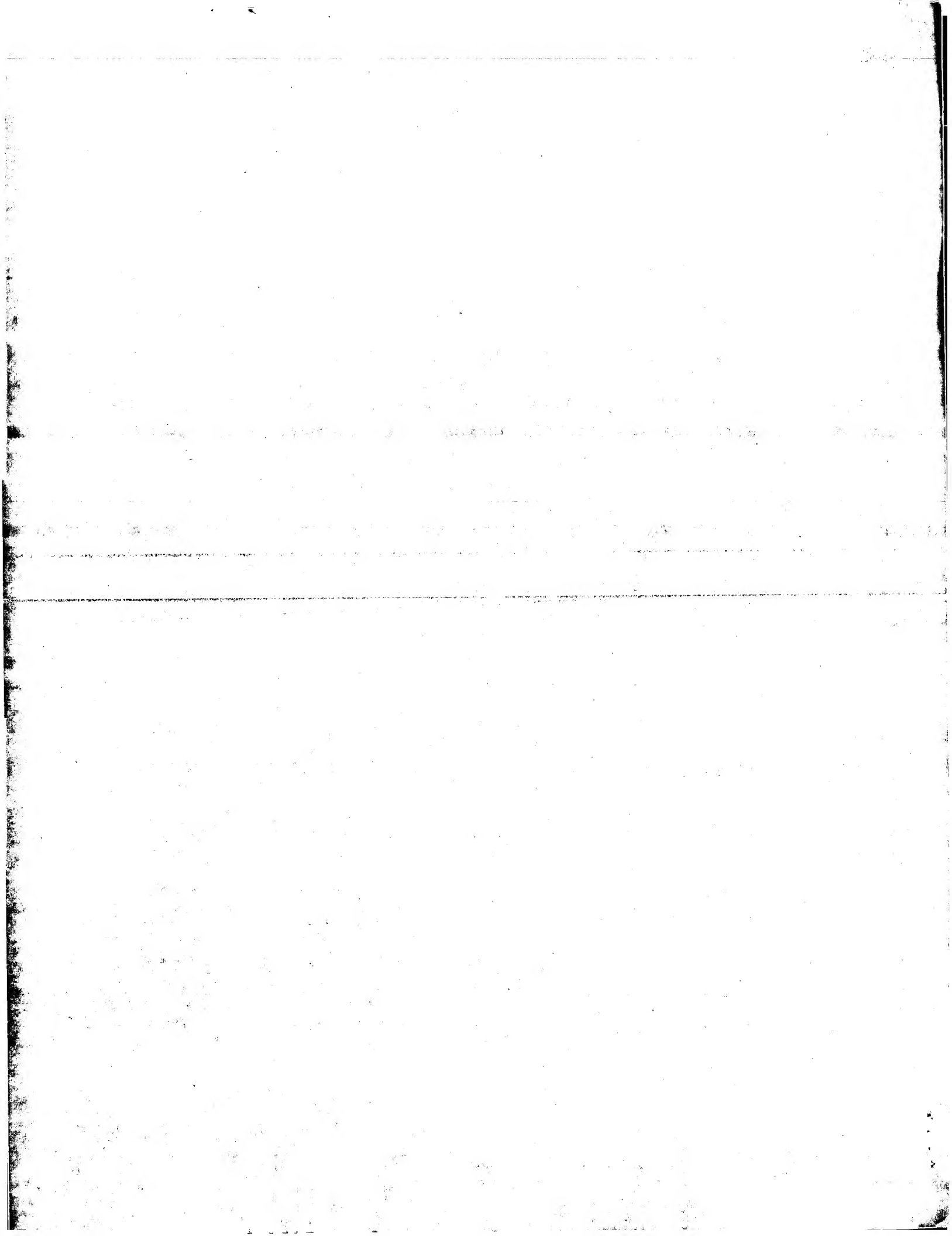
Query Match 7.3%; Score 103.6; DB 2; Length 5952;  
Best Local Similarity 80.7%; Pred. No. 3.7e-23;  
Matches 121; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
Qy 1 tgagcagtactcgttgcgcgcgcgcacacacacataatagctgacagactaacaga 60  
Db 1587 TGAGCAGTACTCGTTGCTGCGCGCGCCACACAGACATAATAGCTGACAGACTAACAGA 1646  
Qy 61 ctgttcctttccatgggtcttttgcagtcacccgtgacacacgtgtgatcagatatc 120  
Db 1647 CTGTTCCCTTCATGGGCTCTTTTGTGAGTCACCGTCGTCGACACAGAGCTGAGATCCTAC 1706  
Qy 121 ccacatgttgccttagggaaacctctg 150  
Db 1707 AGGAGTCCAGGGCTGGAGAGAAACCTCTG 1736

RESULT 14  
US-08-450-945-57/c  
; Sequence 57, Application US/08450945  
; Patent No. 5783383  
; GENERAL INFORMATION:  
; APPLICANT: Kondo, Kazuhiro  
; APPLICANT: Mocarski, Edward S. Jr.  
; TITLE OF INVENTION: LATENT TRANSCRIPTS AND PROMOTERS  
; TITLE OF INVENTION: OF CYTOMEGALOVIRUS











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 21:32:29 ; Search time 7419.31 Seconds  
(without alignments)  
2065.351 Million cell updates/sec

Title: US-09-391-606-12

Perfect score: 1426

Sequence: 1 tggacgtactctgtgtgc.....tcctgacccctggaaggtgc 1426

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703674

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

EST: \*  
1: em\_estfun: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estom: \*  
5: em\_estpl: \*  
6: em\_estba: \*  
7: em\_estro: \*  
8: em\_estov: \*  
9: em\_htc: \*  
10: gb\_est1: \*  
11: gb\_est2: \*  
12: gb\_htc: \*  
13: gb\_gss: \*  
14: em\_gss\_fun: \*  
15: em\_gss\_hum: \*  
16: em\_gss\_inv: \*  
17: em\_gss\_pln: \*  
18: em\_gss\_pro: \*  
19: em\_gss\_rod: \*  
20: em\_gss\_vrt: \*  
21: em\_gss\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70	4.9	149	10	AW669676
2	70	4.9	162	10	AW669522
3	70	4.9	162	10	AW785253
4	70	4.9	162	10	AW785901
5	70	4.9	243	10	AW689518
6	70	4.9	328	10	AW669634
7	70	4.9	374	10	AW669633
8	70	4.9	503	11	BF074689
9	54	3.8	98	11	BG365021
10	42.2	3.0	566	13	AQ778622
11	41.6	2.9	337	11	BF712368
12	41.6	2.9	359	11	BF711512

C 13	41.6	2.9	359	11	BF712352
C 14	41.6	2.9	396	11	BF712329
C 15	41.6	2.9	403	11	BF702314
C 16	41.6	2.9	432	11	BF711501
C 17	41.6	2.9	448	11	BF712292
C 18	41.6	2.9	448	11	BF712400
C 19	41.6	2.9	520	11	BF712343
C 20	41.6	2.9	538	11	BF703911
C 21	41.6	2.9	534	11	BF712419
C 22	41.6	2.9	537	11	BF702273
C 23	40.2	2.8	593	11	BF263367
C 24	40	2.8	143	10	AW359614
C 25	40	2.8	294	11	BF19164
C 26	40	2.8	294	11	BF19165
C 27	40	2.8	358	11	BF711604
C 28	40	2.8	359	11	BF703303
C 29	40	2.8	363	11	BF702121
C 30	40	2.8	404	11	BF711339
C 31	40	2.8	432	11	BF712300
C 32	40	2.8	433	11	BG894486
C 33	40	2.8	439	11	BF712448
C 34	40	2.8	458	11	BF712424
C 35	40	2.8	526	11	BF702026
C 36	40	2.8	527	11	BF701994
C 37	40	2.8	532	11	BF702078
C 38	40	2.8	535	11	BF702139
C 39	40	2.8	537	11	BF702127
C 40	40	2.8	575	11	BF712372
C 41	35.8	2.8	582	11	BF198279
C 42	35.2	2.7	1101	13	CNS0039G
C 43	39	2.7	675	13	AZ840016
C 44	39	2.7	690	10	AU034107
C 45	38.4	2.7	1321	11	BG490965

## ALIGNMENTS

## RESULT

AW669676  
LOCUS 149 bp mRNA EST 25-APR-2001  
DEFINITION 113173 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION AW669676  
VERSION AW669676.1 GI:7526190  
KEYWORDS EST.  
SOURCE BOS taurus  
COW.

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

## REFERENCE

1 (bases 1 to 149)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W., Rohrer,G.A., Chitko-McKown,C.G., Perteau,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.

## TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

## JOURNAL

Genome Res. 11 (4), 626-630 (2001)

## MEDLINE

21180013

## COMMENT

Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smithemail@marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCACGCG  
Plate: 107 row: D column: 22



Seq primer: ATTAGGTGACACTATAG.

## FEATURES

source

Location/Qualifiers

1. 149

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 1BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;

Library made from pooled tissue from lymph node, ovary,

fat, hypothalamus, and pituitary."

24 a 48 c 38 g 39 t

## BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 7.5e-10; Length 149;

Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1357 ctgtgccttctagtgcagccatctgtgttgcctccctcccggtgccttccttgacc 1416

Db 36 CTGTGCTTCTAGTGCAGCCATCTGTGTTCGCCCTCCCGGTGCTTCTCTTGACCC 95

QY 1417 tgaaaggtgc 1426

Db 96 TGAAGGTGC 105

## RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

Location/Qualifiers

1. 162

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 1BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;

Library made from pooled tissue from lymph node, ovary,

fat, hypothalamus, and pituitary."

28 a 51 c 42 g 41 t

## BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 7.8e-10; Length 162;

Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1357 ctgtgccttctagtgcagccatctgtgttgcctccctcccggtgccttccttgacc 1416

Db 49 CTGTGCTTCTAGTGCAGCCATCTGTGTTCGCCCTCCCGGTGCTTCTCTTGACCC 108

QY 1417 tgaaaggtgc 1426

Db 109 TGAAGGTGC 118

## RESULT 3

AW785253

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. 162

/organism="Sus scrofa"

/db\_xref="taxon:9823"

/clone\_lib="MARC 1PIG"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;

Library made from pooled tissue from day 11, 13, 15, 20,

and 30 embryos."

28 a 51 c 42 g 41 t

## BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 7.8e-10; Length 162;

Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1357 ctgtgccttctagtgcagccatctgtgttgcctccctcccggtgccttccttgacc 1416

Db 49 CTGTGCTTCTAGTGCAGCCATCTGTGTTCGCCCTCCCGGTGCTTCTCTTGACCC 108

QY 1417 tgaaaggtgc 1426



|||||||

Db 109 TGAAGGTGC 118

RESULT 4

AW785901

LOCUS 162 bp mRNA EST 09-JUL-2000

DEFINITION 117399 MARC lPig Sus scrofa cDNA 5', mRNA sequence.

ACCESSION AW785901

VERSION AW785901.1 GI:7842677

KEYWORDS EST.

SOURCE pig.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 162)

AUTHORS Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W., and Keele, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

JOURNAL Unpublished (2000)

COMMENT Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGCG

Plate: 37 row: K column: 23

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1..162

/organism="Sus scrofa"

/db\_xref="taxon:9823"

/clone\_lib="MARC lPig"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;

Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT 28 a 51 c 42 g 41 t

ORIGIN

Query Match 4.9%; Score 70; DB 10; Length 162;

Best Local Similarity 100.0%; Pred. No. 7.8e-10;

Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1357 cgtgccccttctagttgcagccatctgtttgttgcctccccccttgccttcttgcaccc 1416

|||||||

Db 49 cgtgccccttctagttgcagccatctgtttgttgcctccccccttgccttcttgcaccc 108

|||||||

QY 1417 tggaaaggtgc 1426

|||||||

Db 109 TGAAGGTGC 118

RESULT 5

AW669518

LOCUS 243 bp mRNA EST 25-APR-2001

DEFINITION 112921 MARC lBOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION AW669518

VERSION AW669518.1 GI:7526032

KEYWORDS EST.

SOURCE cow.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 328)

AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keele, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 243)

AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keele, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

CONTACT: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGCG

Plate: 106 row: J column: 22

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1..243

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC lBOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;

Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

BASE COUNT 56 a 67 c 59 g 61 t

ORIGIN

Query Match 4.9%; Score 70; DB 10; Length 243;

Best Local Similarity 100.0%; Pred. No. 9e-10;

Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1357 cgtgccccttctagttgcagccatctgtttgttgcctccccccttgccttcttgcaccc 1416

|||||||

Db 49 cgtgccccttctagttgcagccatctgtttgttgcctccccccttgccttcttgcaccc 108

|||||||

QY 1417 tggaaaggtgc 1426

|||||||

Db 109 TGAAGGTGC 118

RESULT 6

AW669634

LOCUS 328 bp mRNA EST 25-APR-2001

DEFINITION 113105 MARC lBOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION AW669634

VERSION AW669634.1 GI:7526148

KEYWORDS EST.

SOURCE cow.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 328)

AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keele, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.



Genome Res. 11 (4), 626-630 (2001)

21180013

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -mnscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCCGAGTCAGCAGC

Plate: 107 row: I column: 14

Seq primer: ATTTAGGTGACACTATAG.

Location/Qualifiers

1. .328

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC LBOV"

/tissue\_type="pooled"

/lab\_host="DH108"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;

Library made from pooled tissue from lymph node, ovary,

fat, hypothalamus, and pituitary."

71 a 88 c 101 g 68 t

BASE COUNT

ORIGIN

Query Match 4.9% Score 70; DB 10; Length 328;

Best Local Similarity 100.0%; Pred. No. 1e-09; 0; Indels 0; Gaps 0;

Matches 70; Conservative 0; Mismatches 0;

QY 1357 ctgtgccttctagtgcagccatctgtgtgtgccccctccccgcgttccttcgtgaccc 1416

Db 110 CTGTGCTTCTAGTTGTCAGCACCATCTGTGTGTCCTCCCGTCCTTCCTTGACCC 51

QY 1417 tggaaagtgc 1426

Db 50 TGGAAAGTGC 41

RESULT 7

AW669633/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 374)

Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,

Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett

G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,

Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and

Keefe,J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA

libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

21180013

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred

v0.980904.e. Vector identified by cross\_match with the -mnscore 18

```

and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 107 row: I column: 13
Seq primer: ATTTAGGTGACACTATAG.
FEATURES             source
Location/Qualifiers
1..374
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
79 t
BASE COUNT      84 a 102 c 109 g
ORIGIN
Query Match      4.9%; Score 70; DB 10; Length 374;
Best Local Similarity 100.0%; Pred. No. 1e-09; 0; Gaps
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1357 ctgtgcctctagtgcagcccatctgtttgtgcccctccccgcgtcctcttgacc 1416
DB 109 CTGTGCTTCTAGTGGCCAGCATCTGTCTTGCCCTCCCGCTCCTTGCCTGACCC 50
QY 1417 tggaaaggtgc 1426
DB 49 TGGAAAGGTGC 40
RESULT 8
BF074689          503 bp      mRNA          EST          25-APR-2001
LOCUS             222166 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION
ACCESSION         BF074689
VERSION           BF074689.1 GI:10868200
KEYWORDS          EST.
SOURCE            cow.
ORGANISM          Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 503)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
: TITLE           Sequence evaluation of four pooled-tissue normalized bovine cDNA
:                 libraries and construction of a gene index for cattle
: JOURNAL          Genome Res. 11 (4), 626-630 (2001)
: MEDLINE          21180013
: COMMENT          Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.90904.e. Vector identified by cross_match with the -minscore 11
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 82 row: E column: 4
Seq primer: ATTTAGGTGACACTATAG.
FEATURES             source
Location/Qualifiers
1..503
/organism="Bos taurus"
/db_xref="taxon:9913"

```



```

Query Match      3.8%; Score 54; DB 11; Length 98;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 csagccatctgtgtttgtgcccccccccgtccttccctgacctggaaggtgc 1426
db 1 CCAGCCATCTGTGTTTGTGCCCCCCCCCGCGCTTCCTTCACCTCGCAAGGTGC 54

```

RESULT	10
AQ778622/c	
LOCUS	AQ778622 566 bp DNA
DEFINITION	HS_2235_A2_B08_T7C Approved Human Genomic Sperm Library D Homo sapiens clone Plate=2235 Col=16 Row=C, DNA sequence.
ACCESSION	AQ778622
VERSION	AQ778622.1 GI:5681582
KEYWORDS	GSS.
SOURCE	human.

REFERENCE  
1 (bases 1 to 566)

AUTHORS  
Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D. and Hood, L.

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: <a href="http://www.htsc.washington.edu">http://www.htsc.washington.edu</a> Plate: 2235 row: C column: 16 Seq primer: T7 Class: RAC ends

```

FEATURES
source
High quality sequence stop: 566.
Location/Qualifiers
1..566
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=2235 Col=16 Row=C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
160 a 101 c 101 g 201 t 3 others
BASE COUNT
ORIGIN

```

Query Match	3.0%	Score 42.2;	DB 13;	Length 566;
Best Local Similarity	51.6%;	Pred. NO. 0.21;		
Matches	95; Conservative	0; Mismatches	89; Indels	0; Gaps
QY	596	tgtggaactttacagacacacctcttctcttgagcgttaggcgctcgtagagcccttatg	655	
Dd	248	TTTAAATATTACAGAACTACAATTTTTTCCTCAAAGTAATAATCTTATGGAGGTTAAG	189	
QY	656	ggatcggttgtgcaactcttggagcgtgaattccaatatgcacagtccccaaacctaaagt	715	
Dd	188	AGTATGTGAAGATGAAACTAAGCTGTATATGTACTAACACGCACAAGATCAGACATAGTT	129	
QY	716	tgnagaacttaatgtgatctgttaacgtatgcgaattctctgtaaacaaaaaccagggcta	775	
Dd	128	CTAACAAATAACTGCAATATGGTGACTTAGTGCAACTATAGTGAANTTAAAGCCCTATATANA	69	
QY	776	taaa	779	



Db 68 TAAA 65  
||||

## RESULT 11

BF712368/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BF712368 337 bp mRNA EST 02-JAN-2001  
MI-P-A3-adv-d-06-1-UM.s1 MI-P-A3 Sus scrofa cDNA clone  
MI-P-A3-adv-d-06-1-UM 3', mRNA sequence.  
BF712368 1 GI:12011843  
EST.  
SOURCE  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
1 (bases 1 to 337)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477  
Contact: Tuggle CK  
Molecular Genetics Laboratory, Department of Animal Science  
Iowa State University  
201 Kildee Hall, Ames, IA 50011-3150, USA  
Tel: 5152944252  
Fax: 5152942401  
Email: cktuggle@iastate.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
non-normalized anterior pituitary at estrus day 12 library cDNA  
Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal  
Science Research Center, Department of Animal Science, University  
of Missouri-Columbia, 65211 Clone distribution: clones will be  
available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

## FEATURES

Source

Location/Qualifiers

1..337  
/organism="Sus scrofa"  
/strain="crossbreed"  
/db\_xref="taxon:9823"  
/clone="MI-P-A3-adv-d-06-1-UM"  
/clone\_lib="MI-P-A3"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: EcoRI; The MI-P-A3  
library is derived from anterior pituitary at estrus day  
12. For a detailed description of the library from which  
this clone was derived, please visit our web site at  
http://pigest.genome.iastate.edu/  
TAG\_LIB=MI-P-A3  
TAG\_TISSUE=anterior pituitary at estrus day 12  
TAG\_SEQ=TAGACAG"

BASE COUNT

ORIGIN

70 a 81 c 106 g 80 t

Query Match

Best Local Similarity

Matches

2.9%; Score 41.6; DB 11; Length 337;

59; Conservative

0; Mismatches

9; Indels

2; Gaps

1;

QY 1357 ctgtgccttctagttgcagccatctgtgttgcctcccccgtgccttcttgcacc 1416

Db 133 CTGTGCTTCTAGTTGCTGGCATCTCT--GTTGCCCTCCCGCAGTACCTCCCTTGACCC 76

QY 1417 tggaaagtgc 1426

Db 75 TGGAAAGTGC 66

## RESULT 12

BF711512/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BF711512 359 bp mRNA EST 02-JAN-2001  
MI-P-A2-acq-e-09-1-UM.s1 MI-P-A2 Sus scrofa cDNA clone  
MI-P-A2-acq-e-09-1-UM 3', mRNA sequence.  
BF711512 1 GI:12010987  
EST.  
SOURCE  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
1 (bases 1 to 359)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477  
Contact: Tuggle CK  
Molecular Genetics Laboratory, Department of Animal Science  
Iowa State University  
201 Kildee Hall, Ames, IA 50011-3150, USA  
Tel: 5152944252  
Fax: 5152942401  
Email: cktuggle@iastate.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
non-normalized anterior pituitary at estrus day 5 library cDNA  
Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal  
Science Research Center, Department of Animal Science, University  
of Missouri-Columbia, 65211 Clone distribution: clones will be  
available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

## FEATURES

Source

Location/Qualifiers

1..359  
/organism="Sus scrofa"  
/strain="crossbreed"  
/db\_xref="taxon:9823"  
/clone="MI-P-A2-acq-e-09-1-UM"  
/clone\_lib="MI-P-A2"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: EcoRI; The MI-P-A2  
library is derived from anterior pituitary at estrus day  
5. For a detailed description of the library from which  
this clone was derived, please visit our web site at  
http://pigest.genome.iastate.edu/  
TAG\_LIB=MI-P-A2  
TAG\_TISSUE=anterior pituitary at estrus day 5  
TAG\_SEQ=TCGCGT"

BASE COUNT

ORIGIN

68 a 91 c 108 g 92 t

Query Match

Best Local Similarity

Matches

2.9%; Score 41.6; DB 11; Length 359;

59; Conservative

0; Mismatches

9; Indels

2; Gaps

1;

QY 1357 ctgtgccttctagttgcagccatctgtgttgcctcccccgtgccttcttgcacc 1416

Db 133 CTGTGCTTCTAGTTGCTGGCATCTCT--GTTGCCCTCCCGCAGTACCTCCCTTGACCC 76

QY 1417 tggaaagtgc 1426

Db 75 TGGAAAGTGC 66



RESULT 13  
BF712352/c

LOCUS  
DEFINITION MI-P-A3-adf-b-07-1-UM.sl MI-P-A3 Sus scrofa cDNA clone EST 02-JAN-2001  
MI-P-A3-adf-b-07-1-UM 3', mRNA sequence.

ACCESSION  
VERSION BF712352  
KEYWORDS  
SOURCE EST

BF712352.1 GI:12011827

pig.

ORGANISM  
Sus scrofa

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

TITLE  
1 (bases 1 to 359)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Tuggle CK

Molecular Genetics Laboratory, Department of Animal Science

Iowa State University

201 Kildee Hall, Ames, IA 50011-3150, USA

Tel: 5152944252

Fax: 5152942401

Email: cktuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to verify it as a clone from the

non-normalized anterior pituitary at estrus day 12 library cDNA

Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal

Science Research Center, Department of Animal Science, University

of Missouri-Columbia, 65211 Clone distribution: clones will be

available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=yes.

FEATURES  
source

Location/Qualifiers

1..359

/organism="Sus scrofa"

/strain="crossbreed"

/db\_xref="taxon:9823"

/clone\_lib="MI-P-A3-adf-b-07-1-UM"

/clone\_lib="MI-P-A3"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: EcoRI; The MI-P-A3

library is derived from anterior pituitary at estrus day

12. For a detailed description of the library from which

this clone was derived, please visit our web site at

http://pigest.genome.iastate.edu/.

TAG\_LIB=MI-P-A3

TAG\_TISSUE=anterior pituitary at estrus day 12

TAG\_SEQ=TGACAG"

BASE COUNT 71 a 91 c 106 g 91 t

ORIGIN

Query Match 2.9%; Score 41.6; DB 11; Length 359;

Best Local Similarity 84.3%; Pred No. 0.27; Indels 2; Gaps 1;

Matches 59; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

QY 1357 cgtgctcttagtgccagccatctgtgttggccctcccccgccttccttgaccc 1416

|||||

Db 133 CTGTGCTTCTAGTCTGGGCATCTCT--GTTGCCCTCCCGAGTACCTCCTTGACCC 76

|||||

QY 1417 tggaaagtgc 1426

|||||

Db 75 TGGAAAGTGC 66

RESULT 14

BF712329/c

LOCUS

DEFINITION MI-P-A2-aar-f-03-1-UM.sl MI-P-A2 Sus scrofa cDNA clone EST 22-DEC-2000

3F702314 403 bp mRNA

MI-P-A2-aar-f-03-1-UM 3', mRNA sequence.

BF712329

ACCESSION

VERSION BF712329

KEYWORDS

SOURCE EST

BF712329.1 GI:12011804

pig.

ORGANISM

Sus scrofa

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

LOCUS

DEFINITION

MI-P-A3-adf-b-04-1-UM.sl MI-P-A3 Sus scrofa cDNA clone EST 02-JAN-2001

MI-P-A3-adf-b-04-1-UM 3', mRNA sequence.

BF712329

ACCESSION

VERSION BF712329

KEYWORDS

SOURCE EST

BF712329.1 GI:12011804

pig.

ORGANISM

Sus scrofa

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 396)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Tuggle CK

Molecular Genetics Laboratory, Department of Animal Science

Iowa State University

201 Kildee Hall, Ames, IA 50011-3150, USA

Tel: 5152944252

Fax: 5152942401

Email: cktuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to verify it as a clone from the

non-normalized anterior pituitary at estrus day 12 library cDNA

Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal

Science Research Center, Department of Animal Science, University

of Missouri-Columbia, 65211 Clone distribution: clones will be

available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=yes.

FEATURES  
source

Location/Qualifiers

1..396

/organism="Sus scrofa"

/strain="crossbreed"

/db\_xref="taxon:9823"

/clone\_lib="MI-P-A3-adf-b-04-1-UM"

/clone\_lib="MI-P-A3"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: EcoRI; The MI-P-A3

library is derived from anterior pituitary at estrus day

12. For a detailed description of the library from which

this clone was derived, please visit our web site at

http://pigest.genome.iastate.edu/.

TAG\_LIB=MI-P-A3

TAG\_TISSUE=anterior pituitary at estrus day 12

TAG\_SEQ=TGACAG"

BASE COUNT 77 a 103 c 118 g 98 t

ORIGIN

Query Match 2.9%; Score 41.6; DB 11; Length 396;

Best Local Similarity 84.3%; Pred. No. 0.28; Indels 2; Gaps 1;

Matches 59; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

QY 1357 cgtgctcttagtgccagccatctgtgttggccctcccccgccttccttgaccc 1416

|||||

Db 133 CTGTGCTTCTAGTCTGGGCATCTCT--GTTGCCCTCCCGAGTACCTCCTTGACCC 76

|||||

QY 1417 tggaaagtgc 1426

|||||

Db 75 TGGAAAGTGC 66

RESULT 15

BF702314/c

LOCUS

DEFINITION MI-P-A2-aar-f-03-1-UM.sl MI-P-A2 Sus scrofa cDNA clone EST 22-DEC-2000

3F702314 403 bp mRNA

MI-P-A2-aar-f-03-1-UM 3', mRNA sequence.

BF702314

ACCESSION

VERSION BF702314

KEYWORDS

SOURCE EST

BF702314



Search completed: February 7, 2002, 21:32:32  
Job time: 20666 sec



DE XX Human LYST1 longer isoform.

KW XX LYST1; human; lysosomal trafficking regulator;

KW XX Chediak-Higashi syndrome; CH syndrome; autoimmune disease; tumour;

KW XX asthma; urticaria; inflammatory bowel disease; psoriasis;

KW XX systemic lupus erythematosus; rheumatoid arthritis;

KW XX glomerulonephritis; multiple sclerosis;

KW XX post-angioplasty restenosis; vaccine; therapy; diagnosis.

OS XX Homo sapiens.

PN XX WO9728262-A1.

XX XX 07-AUG-1997.

XX XX 31-JAN-1997; 97WO-US01748.

XX XX 23-DEC-1996; 96US-0034346.

PR 01-FEB-1996; 96US-0011146.

PR 20-DEC-1996; 96US-0033599.

XX XX (UYFL ) UNIV FLORIDA.

XX XX Barbosa-Alleyne MDFS, Kingsmore SF;

XX XX WPI; 1997-402616/37.

DR N-PSDB; AAT74201.

XX XX Mammalian lysosomal trafficking regulators LYST1, LYST2 and LYST2 - useful to diagnose Chediak-Higashi syndrome

PS Claim 3; Page 108-109; 237pp; English.

XX XX This protein sequence comprises a lysosomal trafficking regulator (LYST1) polypeptide (long isoform) that is encoded by a gene that is mutated in several Chediak-Higashi syndrome (CHS) patients. CC LYST1 long isoform can be expressed in host cells using an isolated CC LYST1 gene (see AAT74201) for use in various pharmacological and CC immunological applications. A shorter LYST1 isoform (see AAT23597), CC resulting from alternative splicing, has also been identified. CC LYST1 regulates degranulation of lysosomes, late endosomes and CC acidic secretory granules, primarily in leukocytes. Inhibition of CC such degranulation using dominant negatively acting truncated LYST1 CC peptides may be used to treat inflammatory and autoimmune diseases CC e.g. asthma, urticaria, inflammatory bowel disease, systemic lupus CC erythematosus, rheumatoid arthritis, psoriasis, systemic vasculitis, CC glomerulonephritis, multiple sclerosis and post-angioplasty CC restenosis, while LYST1 peptides that mimic or augment LYST1 CC function may be used to treat neoplasia.

XX XX Sequence 2001 AA;

Query Match 6.6%; Score 77.5; DB 18; Length 2001;  
Best Local Similarity 22.2%; Pred. No. 56;  
Matches 41; Conservative 15; Mismatches 50; Indels 79; Gaps 7;

QY 31 NTFVPSSETTEKATNAMYKYCVQWLVGKHSQVP-----WINGOKKPLYL 78  
Db 1585 niflpsk-----wqlvltlylqpggkrrrhgkisiwsggkpdvt 1626

QY 79 YGAFLMPLAKATKTTLNGKENLAW-FITGTLGG-----LRKAGDWSATVRYEYVLSVP 133  
27 ldfml-----prktslssdnkfcfmighclssqeeqlglagkwdl-----1667

EIDVSGIGRGNLLKFWFAQAATANYDPKEANSFTNYKGFSAIYMGITDLSLFRAYGAYS 193  
-----gnllllfngakvsg-----eaflyacgphntsympckyg 1702

198 1707

Fri Feb

RESULT 14  
AAW31950  
ID AAW31950 standard; Protein; 3672 AA.  
XX AAW31950;  
AC AAW31950;  
XX 14-APR-1998 (first entry)  
DT Human bg protein associated with Chediak-Higashi syndrome.  
DE Chediak-Higashi syndrome; beige; bg gene; intracellular vesicle;  
XX human; therapy; diagnosis.  
KW Homo sapiens.  
OS Homo sapiens.  
XX Key Location/Qualifiers  
FH Peptide 3565..3579  
FT /note= "WD40 or G protein-beta subunit repeat motif"  
FT WO9734914-A1.  
XX 25-SEP-1997.  
XX 21-MAR-1997; 97WO-US05068.  
XX 01-JUL-1996; 96US-0021064.  
PR 22-MAR-1996; 96US-0013883.  
PR 19-APR-1996; 96US-0015673.  
XX (MILL-) MILLENNIUM PHARM INC.  
PA (UTAH ) UNIV UTAH.  
XX Kaplan J, Moore KJ, Perou CM;  
XX WPI; 1997-502728/46.  
XX N-PSDB; AAT89257.  
XX Murine and human long and short form beige genes - useful to screen for compounds to treat, or to diagnose intracellular vesicle disorders, specifically Chediak-Higashi syndrome  
XX Claim 9; Fig 8; 133pp; English.  
XX This polypeptide is encoded by an alternatively spliced transcript of the human beige (bg) gene (see AAT89257). The amino acid sequence of the predicted bg gene product indicates that the protein is novel. A longer, putatively full-length bg transcript (see AAT89256) encodes a 3801-amino acid protein (see AAW31949). The bg gene product is involved in the normal differentiation and/or function of intracellular vesicles. Mutation of the human bg gene is responsible for human Chediak-Higashi syndrome (CHS). An intracellular vesicle disorder, specifically CHS, can be diagnosed by measuring bg gene expression in a patient sample, e.g. by detecting mRNA transcripts of the bg gene, the bg gene product or a bg gene mutation contained in the genome of the mammal, specifically in a splice site of the bg gene (claimed). Genetically engineered host cells can be used to screen for compounds useful for the treatment of intracellular vesicle disorders, specifically CHS, e.g. a molecule that modulates the expression of the bg gene in a mammal (claimed).

XX Sequence 3672 AA;

Query Match 6.6%; Score 77.5; DB 18; Length 3672;  
Best Local Similarity 22.2%; Pred. No. 1.3e+02;  
Matches 41; Conservative 15; Mismatches 50; Indels 79; Gaps 7;

QY 31 NTFVPSSETTEKATNAMYKYCVQWLVGKHSQVP-----WINGOKKPLYL 78  
Db 1585 niflpsk-----wqlvltlylqpggkrrrhgkisiwsggkpdvt 1626



DE H. pylori outer membrane polypeptide.  
 XX Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;  
 KW cellular immune response.  
 XX Helicobacter pylori.  
 OS WO9921959-A2.  
 XX 06-MAY-1999.  
 XX 28-OCT-1998; 98WO-US22883.  
 XX 17-DEC-1997; 97US-0993001.  
 XX 28-OCT-1997; 97US-0959131.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 XX Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;  
 PI WPI; 1999-326698/27.  
 DR N-PSDB; AAX75836.  
 XX Cellular vaccine against Helicobacter pylori  
 PT Claim 7; Page 311-312; 352pp; English.  
 XX The invention relates to a vaccine for preventing or treating infections  
 CC by Helicobacter pylori. The vaccine contains at least one isolated  
 CC H. pylori polypeptide, or its fragments, in a carrier, where the  
 CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a  
 CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce  
 CC humoral and cellular immune responses. The vaccines are used to treat or  
 CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent  
 CC nucleic acid sequences encoding H. pylori outer membrane polypeptides  
 CC (OMPs) AAX17160 to AAX17218.  
 XX Sequence 327 AA;  
 SQ

Query Match 6.8%; Score 79; DB 20; Length 327;  
 Best Local Similarity 25.7%; Pred. No. 3;  
 Matches 58; Conservative 23; Mismatches 71; Indels 74; Gaps 17;

QY 48 AMKYKVC-VQWLKGVKHSOVPHN-GQKKPLYLYGAFLLMPLAKATKTTLNGKENLAWFI 105  
 DB 118 amtnyvgwggilvg---qkpwscglatgnltgggy--dkmtqgmtqisnqeala--- 169  
 QY 106 GTTLG-----GLRKAGDWSATVRYEVEALSV-----PEIDVSGIGRG-----NL--L 146  
 DB 170 astcaktadyadhrnyviynayirnykdifeirgryespadymsgynqgldmtlnlgnf 229  
 QY 147 KFW----FAQAJAAN-----YDPKE---ANSFT-----NYKGSALYMYGIT 181  
 DB 230 kfwfssfgfaynewlynfyspktytlknqgtinpgvhafiiwnykgfs-----iq 283  
 QY 182 DLSLFRAYGAY-----SPANDKLGSDFTFR-KFDLGIISAF 217  
 DB 284 pfvyrspfnedpnftdydsnptftglg----frsqtdvtvlnpf 325

RESULT 12  
 AAY96159  
 ID AAY96159 standard; Protein; 753 AA.  
 XX AAY96159;  
 XX 19-DEC-2000 (first entry)  
 XX Arabidopsis thaliana OPT family member emb CAB43855.1.  
 XX Oligopeptide transporter; OPT; isp4-like protein; OPT1; enkephalin;  
 KW oploid.

XX Arabidopsis thaliana.  
 OS WO200052162-A2.  
 XX 08-SEP-2000.  
 XX 01-MAR-2000; 2000WO-US05158.  
 XX 01-MAR-1999; 99US-0122312.  
 XX (UYTE-) UNIV TENNESSEE RES CORP.  
 XX (BECK/) BECKER J M.  
 XX (HAUS/) HAUSER M.  
 XX (DONH/) DONHARDT A.  
 XX (BARN/) BARNES D.  
 XX Becker JM, Hauser M, Donhardt A, Barnes D;  
 PI WPI; 2000-587311/55.  
 XX Obtaining mammalian enkephalin transport proteins for transporting  
 PT enkephalin across cell membrane, comprises transforming OPT1 gene  
 PT deleted yeast strain with mammalian gene library and selecting the  
 PT cells -  
 XX Disclosure; Fig 5; 59pp; English.  
 XX The present sequence is that of an oligopeptide transporter (OPT)  
 CC family member (isp4-like protein) from Arabidopsis thaliana,  
 CC designated emb CAB43855.1. Members of the OPT family have been  
 CC identified in yeasts and plants. The invention relates to a novel  
 CC OPT of Saccharomyces cerevisiae, designated OPT1 (see AAY96149).  
 CC OPT1 is the first example of a genetically defined eukaryotic  
 CC transport protein which can transport enkephalins across the cell  
 CC membrane. Vectors allowing expression of the OPT1 gene, transgenic  
 CC plants, and treatment of OPT1p with toxic enkephalins as an  
 CC antifungal method are claimed. Also claimed is a method for  
 CC obtaining mammalian enkephalin transporters by functional  
 CC complementation of OPT1 deficient yeast.  
 XX Sequence 753 AA;  
 SQ

Query Match 6.7%; Score 78; DB 21; Length 753;  
 Best Local Similarity 22.2%; Pred. No. 12;  
 Matches 45; Conservative 16; Mismatches 50; Indels 92; Gaps 10;

QY 8 WV---VEGILN--RLPKQ-----FFVKCSVVDNNTFVPSETSTSTTEKAATNAMYKY 53  
 DB 564 wllssvenicntdmlpksspwtpcpdvvfynasii-wgllpggrmftskgi-----ypg 616  
 QY 54 CVWQWLKGVKHSQVP-----WINGQKKPLYLYGAFLLMPLAKATKTTLNGKENL 101  
 DB 617 mmwffligflapvpwvffarkfpekkwhqihiplifsganvm-pmakavhy----- 667  
 QY 102 AWFIGTGLGLRGKAGDWSATVRYEVEALSVPEIDVSGIGRGNLLKFWFAQAIANYDPK 161  
 DB 668 -----ws-----wfvavgifvny--- 679

QY 162 EANSFTNYKGSALYMYGITDLSL 184  
 DB 680 --yifrykgwwarhnyllsaal 700

RESULT 13  
 AAY23596  
 ID AAY23596 standard; Protein; 2001 AA.  
 XX AAY23596;  
 XX 10-FEB-1998 (first entry)  
 XX



QY 79 YGAFMLNPLAKATKTTLNGKENLAW-FIGGTLGG-----LRKAGDWSATVRYEYVEALSVP 133  
 Db 1627 ldfml-----prktslssdsntfcmighclssqeeflqlagkwdl----- 1667  
 QY 134 EIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALMYGITDSLSFRAYGAYS 193  
 Db 1668 -----gnlllfnagkvgqsq-----eaflyacgpnhtsvmpckyg 1702  
 QY 194 KPAND 198  
 Db 1703 kfvnd 1707

RESULT 15  
 AAW31949  
 ID AAW31949 standard; Protein; 3801 AA.  
 XX AAW31949;  
 AC AAW31949;  
 XX 14-APR-1998 (first entry)  
 XX Human bg protein associated with Chediak-Higashi syndrome.  
 XX Chediak-Higashi syndrome; beige; bg gene; intracellular vesicle;  
 KW human; therapy; diagnosis.  
 XX Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 3694..4708  
 FT /note= "WD40 or G protein-beta subunit repeat motif"  
 FT Misc-difference 3767  
 FT /note= "encoded by TNC"  
 XX  
 PN WO9734914-A1.  
 XX  
 PD 25-SEP-1997.  
 XX  
 PF 21-MAR-1997; 97WO-US05068.  
 XX  
 PR 01-JUL-1996; 96US-0021064.  
 PR 22-MAR-1996; 96US-0013883.  
 PR 19-APR-1996; 96US-0015673.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 PA (UTAH) UNIV UTAH.  
 XX  
 PI Kaplan J, Moore KJ, Perou CM;  
 XX  
 DR WPI; 1997-502728/46.  
 DR N-PSDB; AAT89256.  
 XX  
 PT Murine and human long and short form beige genes - useful to screen  
 PT for compounds to treat, or to diagnose intracellular vesicle  
 PT disorders, specifically Chediak-Higashi syndrome  
 XX  
 PS Claim 9; Fig 7; 133pp; English.  
 XX  
 CC This polypeptide is encoded by the novel human beige (bg) gene  
 CC (see AAT89256) that is involved in the normal differentiation and/or  
 CC function of intracellular vesicles. The amino acid sequence of  
 CC the predicted bg gene product indicates that the protein is novel.  
 CC An alternatively spliced transcript of the human bg gene (see  
 CC AAT89257) encodes a 3672-amino acid protein (see AAW31950). Mutation  
 CC of the human bg gene is responsible for human Chediak-Higashi  
 CC syndrome (CHS). An intracellular vesicle disorder, specifically  
 CC CHS, can be diagnosed by measuring bg gene expression in a patient  
 CC sample, e.g. by detecting mRNA transcripts of the bg gene, the bg  
 CC gene product or a bg gene mutation contained in the genome of the  
 CC mammal, specifically in a splice site of the bg gene (claimed).  
 CC Genetically engineered host cells can be used to screen for

CC compounds useful for the treatment of intracellular vesicle  
 CC disorders, specifically CHS, e.g. a molecule that modulates the  
 CC expression of the bg gene in a mammal (claimed).  
 XX Sequence 3801 AA;  
 SQ

Query Match 6.6%; Score 77.5; DB 18; Length 3801;  
 Best Local Similarity 22.2%; Pred. No. 1.4e+02;  
 Matches 41; Conservative 15; Mismatches 50; Indels 79; Gaps 7;  
 QY 31 NTFVPSETTTEKAATNAMYKYCYWQNLVGHKSQVP-----WINGQKKPLYL 78  
 Db 1585 niflpsk-----wqlvltylqpgqkrrrhgkislwvsgqrkpdvt 1626  
 QY 79 YGAFMLNPLAKATKTTLNGKENLAW-FIGGTLGG-----LRKAGDWSATVRYEYVEALSVP 133  
 Db 1627 ldfml-----prktslssdsntfcmighclssqeeflqlagkwdl----- 1667  
 QY 134 EIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALMYGITDSLSFRAYGAYS 193  
 Db 1668 -----gnlllfnagkvgqsq-----eaflyacgpnhtsvmpckyg 1702  
 QY 194 KPAND 198  
 Db 1703 kfvnd 1707

Search completed: February 7, 2002, 21:34:34  
 Job time: 20753 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:54:39 ; Search time 243.49 Seconds  
(without alignments)  
1024.076 Million cell updates/sec

Title: US-09-391-606-14  
Perfect score: 1101  
Sequence: 1 atgttgcttaggaaccc.....tatctggtcagtcagattc 1101

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37	3.4	40	US-09-025-176-2	Sequence 2, Appli
C 2	35.2	3.2	48	US-08-116-389-6	Sequence 6, Appli
C 3	35.2	3.2	48	US-08-708-431-6	Sequence 6, Appli
C 4	35.2	3.2	48	US-08-880-830-6	Sequence 6, Appli
C 5	35.2	3.2	48	PCT-US94-13895-6	Sequence 6, Appli
C 6	34.2	3.1	1978	US-08-137-175A-4	Sequence 4, Appli
C 7	34.2	3.1	1978	US-08-479-017-4	Sequence 4, Appli
C 8	34.2	3.1	3167	US-08-472-534-1	Sequence 1, Appli
C 9	34.2	3.1	4320	US-08-472-534-4	Sequence 4, Appli
C 10	33.2	3.0	490	US-08-245-511-1	Sequence 1, Appli
C 11	33.2	3.0	490	US-08-600-993A-1	Sequence 1, Appli
C 12	33.2	3.0	2019	US-08-245-511-46	Sequence 46, Appli
C 13	33.2	3.0	2019	US-08-600-993A-46	Sequence 46, Appli
C 14	32.2	2.9	1569	US-08-680-726A-57	Sequence 57, Appli
C 15	32.2	2.9	1569	US-09-092-409-57	Sequence 57, Appli
C 16	32.2	2.9	3263	US-08-714-918-20	Sequence 20, Appli
C 17	32.2	2.9	3263	US-09-265-315-20	Sequence 20, Appli
C 18	32.2	2.9	3263	US-09-265-315-20	Sequence 20, Appli
C 19	32.2	2.9	3263	US-09-266-417-20	Sequence 20, Appli
C 20	32.2	2.9	10592	US-08-680-726A-51	Sequence 51, Appli
C 21	32.2	2.9	10592	US-08-680-726A-52	Sequence 52, Appli
C 22	32.2	2.9	10592	US-09-092-409-51	Sequence 51, Appli
C 23	32.2	2.9	10592	US-09-092-409-52	Sequence 52, Appli
C 24	32	2.9	1017	US-08-714-918-45	Sequence 45, Appli
C 25	32	2.9	1017	US-09-265-315-45	Sequence 45, Appli
C 26	32	2.9	1017	US-09-265-315-45	Sequence 45, Appli
C 27	32	2.9	1017	US-09-266-417-45	Sequence 45, Appli

28 31.8 2.9 1005 4 US-09-446-504-79 Sequence 79, Appli  
29 31.6 2.9 1307 2 US-08-867-030B-6 Sequence 6, Appli  
30 31.6 2.9 1307 5 PCT-US95-06119-6 Sequence 6, Appli  
31 31.2 2.8 2359 1 US-08-426-169-5 Sequence 5, Appli  
32 31.2 2.8 2359 4 US-09-233-813-5 Sequence 5, Appli  
33 31.2 2.8 2359 5 PCT-US95-09470-5 Sequence 5, Appli  
34 31.2 2.8 3758 3 US-08-323-477-1 Sequence 1, Appli  
35 31.2 2.8 9432 1 US-08-277-231A-1 Sequence 1, Appli  
36 31.2 2.8 9432 2 US-08-473-750-4 Sequence 4, Appli  
37 31.2 2.8 9432 2 US-08-477-326-4 Sequence 4, Appli  
38 31 2.8 1717 1 US-08-229-515A-12 Sequence 12, Appli  
39 31 2.8 1717 1 US-08-645-863-12 Sequence 12, Appli  
40 31 2.8 3660 1 US-08-158-232-42 Sequence 42, Appli  
41 31 2.8 3660 1 US-08-611-928-42 Sequence 42, Appli  
42 31 2.8 3660 3 US-09-173-891-42 Sequence 42, Appli  
43 30.8 2.8 618 4 US-08-953-326-22 Sequence 22, Appli  
44 30.4 2.8 1425 3 US-09-009-494-5 Sequence 5, Appli  
45 30.4 2.8 1425 4 US-09-010-232-5 Sequence 5, Appli

## ALIGNMENTS

RESULT 1  
US-09-025-176-2/c  
Sequence 2, Application US/09025176  
Patent No. 6258532  
GENERAL INFORMATION:  
APPLICANT: Charles W. Stratton  
APPLICANT: William M. Mitchell  
TITLE OF INVENTION: Methods for in vitro and in vivo  
susceptibility testing of Chlamydia species  
FILE REFERENCE: 50150/006003  
CURRENT APPLICATION NUMBER: US/09/025,176  
CURRENT FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: 08/911,593  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 40  
TYPE: DNA  
ORGANISM: Chlamydia pneumoniae  
US-09-025-176-2

Query Match 3.4%; Score 37; DB 4; Length 40;  
Best Local Similarity 100.0%; Pred. No. 0.0065;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1065 cgagagagctgcacgtatctggtcagtcagttcagattc 1101  
|||||  
Db 40 CGAGAGAGCTGCTACGCTATCTGTCAGTTCAGATTTC 4

RESULT 2  
US-08-116-389-6  
Sequence 6, Application US/08116389  
Patent No. 5601978  
GENERAL INFORMATION:  
APPLICANT: Burczak, John  
APPLICANT: Carrino, J.J.  
APPLICANT: Salituro, J.A.  
APPLICANT: Pabich, E.K.  
APPLICANT: Klonowski, P.A.  
APPLICANT: Manlove, M.T.  
APPLICANT: Marshall, R.L.  
TITLE OF INVENTION: Materials and Methods for the Detection  
of Chlamydia Trachomatis  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES  
STREET: One Abbott Park Road



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/116,389  
FILING DATE: 03 SEPTEMBER 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Brainard, Thomas D.  
REGISTRATION NUMBER: 32,459  
REFERENCE/DOCKET NUMBER: 5372. US 01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708 937-4884  
TELEFAX: 708 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Chlamydia trachomatis  
IS-08-708-431-6

Query Match	3.28;	Score 35.2;	DB1; Length 48;
Best Local Similarity	83.3%;	Pred. No. 0.028;	
Matches	40; Conservative	0; Mismatches	8; Indels
QY	318	tggggaatgcgtttgatatttctgtacttttagggagccttcataatggta	365
b	1	ttwgggacccentttcatcttatcttcttacatttaggagccaccacttgata	48

RESULT 4  
US-08-880-830-6  
; Sequence 6, Application US/08880830  
; Patent No. 5845785  
; GENERAL INFORMATION:  
; APPLICANT: Burczak, John  
; APPLICANT: Carrino, J.J.  
; APPLICANT: Salituro, J.A.  
; APPLICANT: Padich, E.K.  
; APPLICANT: Klonowski, P.A.  
; APPLICANT: Manlove, M.T.  
; APPLICANT: Marshall, R.L.  
; TITLE OF INVENTION: Materials and Methods for the Detection  
; of Chlamydia Trachomatis  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ABBOTT LABORATORIES  
; STREET: One Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PCDOS/MSDOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/880,830  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/449,863  
; FILING DATE:  
; APPLICATION NUMBER: 08/116,389  
; FILING DATE: 03 SEPTEMBER 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brainard, Thomas D.  
; REGISTRATION NUMBER: 32,459  
; REFERENCE/DOCKET NUMBER: 5372-US-01  
; TELECOMMUNICATION INFORMATION:

CITY: Abbott Park  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PCDOS/MSDOS  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/116,389  
FILING DATE: 03 SEPTEMBER 1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brainard, Thomas D.  
REGISTRATION NUMBER: 32,459  
REFERENCE/DOCKET NUMBER: 5372.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708 937-4884  
TELEFAX: 708 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Chlamydia trachomatis  
US-08-116-389-6

[illegible]

RESULT 3  
US-08-708-431-6  
; Sequence 6, Application US/08708431  
; Patent No. 5756298  
; GENERAL INFORMATION:  
; APPLICANT: Burczak, John  
; APPLICANT: Carrino, J.J.  
; APPLICANT: Salituro, J.A.  
; APPLICANT: Pabich, E.K.  
; APPLICANT: Klonowski, P.A.  
; APPLICANT: Manlove, M.T.  
; APPLICANT: Marshall, R.L.  
; TITLE OF INVENTION: Materials and Methods for the Detection  
; OF INVENTION: of Chlamydia Trachomatis  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ABBOTT LABORATORIES  
; STREET: One Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PCDOS/MSDOS  
; SOFTWARE: WordPerfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/708,431  
; FILING DATE: 05-SEP-1996  
; CLASSIFICATION: 435







Fri Feb 8 08:44:54 2002

NAME/KEY: CDS  
LOCATION: 127..948  
OTHER INFORMATION: /product= "OspA"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 962..1861  
OTHER INFORMATION: /product= "OspB"  
US-08-137-175A-4

Query Match 3.1%; Score 34.2; DB 1; Length 1978;  
Best Local Similarity 48.2%; Pred. No. 0.47;  
Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;  
QY 835 attgtcagcgaacaaactacacagctgttttaaaacttaactgcagcattcgcagcctccttctta 894  
DB 1715 AGTGTGACAGTAAACAACTAAAGATTTTGTGTTCTTAACAGATGGTACCATACAGTA 1774  
QY 895 ctaggaaatgccacagcattgtctactactgattcgttctcagacattcgcagcctccttctta 954  
DB 1775 CAAGCATATGACACAGCAGGTACTAACTTGAGGGCAACTCAAGTGAATTAAGATCTT 1834  
QY 955 tctgtcagatcaacaaagttaataatctagaaaagcttggagttactgttaggagctact 1014  
DB 1835 GCAGCACTTAAAGCTGCTTTAAATAACATAAAAGTAAACATCCTACATCGGCTAATACC 1894  
QY 1015 ttagtgtgctgataaat 1033  
DB 1895 TTTGTAGGTGTTGTTATT 1913

RESULT 7  
US-08-479-017-4  
Sequence 4, Application US/08479017  
Patent No. 6143872  
GENERAL INFORMATION:  
APPLICANT: BARBOUR, Alan G.  
APPLICANT: BERGSTROM, Sven  
APPLICANT: HANSSON, Lennart  
TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDORFERI AND  
TITLE OF INVENTION: PROPHYLAXIS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,017  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/137,175  
FILING DATE: 26-OCT-1993  
APPLICATION NUMBER: PC/US92/08972  
FILING DATE: 22-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, Iver P. 28, 005  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER: BARBOUR-1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1978 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Borrelia burgdorferi  
STRAIN: ACAL  
INDIVIDUAL ISOLATE: Swedish isolate, pt. acrodermatitis  
INDIVIDUAL ISOLATE: chronicum migrans  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 85..104  
OTHER INFORMATION: /function= "Primer"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 111..130  
OTHER INFORMATION: /function= "Primer"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 948..965  
OTHER INFORMATION: /function= "Primer"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 127..948  
OTHER INFORMATION: /product= "OspA"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 962..1861  
OTHER INFORMATION: /product= "OspB"  
US-08-479-017-4

Query Match 3.1%; Score 34.2; DB 3; Length 1978;  
Best Local Similarity 48.2%; Pred. No. 0.47;  
Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;  
QY 835 attgtcagcgaacaaactacacagctgttttaaaacttaactgcagcattcgcagcctccttctta 894  
DB 1715 AGTGTGACAGTAAACAACTAAAGATTTTGTGTTCTTAACAGATGGTACCATACAGTA 1774  
QY 895 ctaggaaatgccacagcattgtctactactgattcgttctcagacattcgcagcctccttctta 954  
DB 1775 CAAGCATATGACACAGCAGGTACTAACTTGAGGGCAACTCAAGTGAATTAAGATCTT 1834  
QY 955 tctgtcagatcaacaaagttaataatctagaaaagcttggagttactgttaggagctact 1014  
DB 1835 GCAGCACTTAAAGCTGCTTTAAATAACATAAAAGTAAACATCCTACATCGGCTAATACC 1894  
QY 1015 ttagtgtgctgataaat 1033  
DB 1895 TTTGTAGGTGTTGTTATT 1913

RESULT 8  
US-08-472-534-1  
Sequence 1, Application US/08472534  
Patent No. 5919620  
GENERAL INFORMATION:  
APPLICANT: Hamel, Josee  
APPLICANT: Brodeur, Bernard R  
APPLICANT: Martin, Denis  
TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM  
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Neave  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10020  
COMPUTER READABLE FORM:



```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,534
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr, James F
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: Biovac-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3167 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: 30..755
FEATURE:
NAME/KEY: CDS
LOCATION: 771..2912
OTHER INFORMATION: /product= "FucI/HSP72 (C-16)
US-08-472-534-1

```

[illegible]

RESULT 9  
US-08-472-534-4  
; Sequence 4, Application US/08472534  
; Patent No. 5919620  
; GENERAL INFORMATION:  
; APPLICANT: Hamel, Josee  
; APPLICANT: Brodeur, Bernard R  
; APPLICANT: Martin, Denis  
; TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM  
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York

```

STATE: New York
COUNTRY: United States of America
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,534
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr, James F
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: Biovac-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4320 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: 3265..4320
OTHER INFORMATION: /product= "NH2-terminal portion of
OTHER INFORMATION: DNA J"
FEATURE:
NAME/KEY: CDS
LOCATION: 682..2502
OTHER INFORMATION: /product= "Heat-Shock Protein 72"
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 682..2502
US-08-472-534-4

Query Match 3.1%; Score 34.2; DB 2; Length 4320;
Best Local Similarity 49.2%; Pred. No. 0.73; Indels 0; Gaps 0;
Matches 90; Conservative 0; Mismatches 93;

QY 259 aataagcattcacgagtcagagtggttcactaatgcaggctcttgacctaaacatt 318
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2504 AAGATGAGTGATTGGATGAAGAGTATCTAAAAATACACGAAAAGTTTATAATGATTTT 2563
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 319 tgggtagctgttaagtgtttctgatactttaggagcttctaattaggttacattagaggaaac 378
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2564 TGTATCAAGCTGATACTATAGACAATCAAAAGATTTTATTGATATATTCCAATAGAA 2623
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 379 tctacagcgttcaatctcgttgtttatttcggaggttaaaggttactactgttaaatgcaaat 438
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2624 TATTGACTAGATATAGAGAAATTATATTAGCTGAGCATGATAGTTCTGTGTCAAAATGAT 2683
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 439 gaa 441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2684 GAA 2686

RESULT 10
US-08-245-511-1
; Sequence 1, Application US/08245511
; Patent No. 5928900
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert

```



APPLICANT: Pearce, Barbara J  
APPLICANT: Tuomanen, Elaine  
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND  
CELLULAR VACCINES BASED THEREON  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,511  
FILING DATE: 18-MAY-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/116,541  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-069 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 490 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
STRAIN: R6  
IMMEDIATE SOURCE:  
CLONE: SPRU98  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..490  
US-08-245-511-1

Query Match 3.0%; Score 33.2; DB 2; Length 490;  
Best Local Similarity 61.6%; Pred. No. 0.45; Mismatches 0; Gaps 0;  
Matches 53; Conservative 0; Indels 33; Indels 0; Gaps 0;  
QY 795 acaatggtctcgagcaacttttgatgataacatccgattgctcagcgaactacc 854  
DB 330 ACAATCCTTGGAGCAACTTTAGGAGCTGATAATGCTATTATTTGATATTTCAACAACACTACA 389  
QY 855 tacagctgttttaaaacttaactgcac 880  
DB 390 AAAAGACGAAGTAACAATATTACAT 415

RESULT 11  
US-08-600-993A-1  
Sequence 1, Application US/08600993A  
Patent No. 5981229  
GENERAL INFORMATION:  
APPLICANT: Masure, H Robert  
APPLICANT: Pearce, Barbara J  
APPLICANT: Tuomanen, Elaine

TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND  
CELLULAR VACCINES BASED THEREON  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/600,993A  
FILING DATE: 1-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/245,511  
FILING DATE: 18-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/116,541  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-069 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 490 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
STRAIN: R6  
IMMEDIATE SOURCE:  
CLONE: SPRU98  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..490  
US-08-600-993A-1

Query Match 3.0%; Score 33.2; DB 2; Length 490;  
Best Local Similarity 61.6%; Pred. No. 0.45; Mismatches 0; Gaps 0;  
Matches 53; Conservative 0; Indels 33; Indels 0; Gaps 0;  
QY 795 acaatggtctcgagcaacttttgatgataacatccgattgctcagcgaactacc 854  
DB 330 ACAATCCTTGGAGCAACTTTAGGAGCTGATAATGCTATTATTTGATATTTCAACAACACTACA 389  
QY 855 tacagctgttttaaaacttaactgcac 880  
DB 390 AAAAGACGAAGTAACAATATTACAT 415

RESULT 12  
US-08-245-511-46  
Sequence 46, Application US/08245511  
Patent No. 5928900  
GENERAL INFORMATION:  
APPLICANT: Masure, H Robert







APPLICANT: Haanes, Elizabeth J.  
APPLICANT: Frank, Rexann S.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/092,409  
FILING DATE: 12-JUL-1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/680,726  
FILING DATE: 12-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1569 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1566  
US-09-092-409-57

Query Match 2.9%; Score 32.2; DB 3; Length 1569;  
Best Local Similarity 43.7%; Pred. No. 1.8; Indels 0; Gaps 0;  
Matches 142; Conservative 0; Mismatches 183;

QY 140 atgttttcgacggtatctttaaagtagatgcacacatttctctatggagcgaagc 199  
DB 107 ATGTTTATAGTACTACCCACCATTAAATGAGACTAAAAATTTACTAGAGGATGGGTAT 166  
QY 200 ctactggatccgctgctgcaaaactatactactgcgtagatagactaacccggcctaca 259  
DB 167 TTTTACAAACCCGCCCTTCTTATTGTAAGATGGGATATCAATATCTATATATGCAATG 226  
QY 260 ataagcatttacacgtagcagagtggttcaactaagcaggttcattgcttaaacattt 319  
DB 227 AACGTAATATTTCTGAAGAAGATATTTTGAATAAACGATGTACAAATTAACACTATTA 286  
QY 320 gggatcgcttgatgtttctgtacttttagggagcctcttaagttacattagggaaact 379  
DB 287 ATTATCCCTAGCTAGCAGATTTTGAGATTAGTAAATTAATACTATTAATAAATAATG 346  
QY 380 ctacagcggtcaactcgttgggtttatccggaggttaaggtactactgtaaatgcaaatg 439  
DB 347 ATGTTTATTTTGTAAATGATAGTGTTCCTCAATAATACTACATAAATAAAGTGTATCC 406  
QY 440 aactaccacaaacgtttcttaagtaa 464  
DB 407 ATATCACAAATGTGACTATATAATAA 431

APPLICANT: Frank, Rexann S.  
APPLICANT: Frank, Rexann S.  
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680,726A  
FILING DATE: 12-JUL-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-46-C1  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1569 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1566  
US-08-680-726A-57

Query Match 2.9%; Score 32.2; DB 1; Length 1569;  
Best Local Similarity 43.7%; Pred. No. 1.8; Indels 0; Gaps 0;  
Matches 142; Conservative 0; Mismatches 183;

QY 140 atgttttcgacggtatctttaaagtagatgcacacatttctctatggagcgaagc 199  
DB 107 ATGTTTATAGTACTACCCACCATTAAATGAGACTAAAAATTTACTAGAGGATGGGTAT 166  
QY 200 ctactggatccgctgctgcaaaactatactactgcgtagatagactaacccggcctaca 259  
DB 167 TTTTACAAACCCGCCCTTCTTATTGTAAGATGGGATATCAATATCTATATATGCAATG 226  
QY 260 ataagcatttacacgtagcagagtggttcaactaagcaggttcattgctttaaacttt 319  
DB 227 AACGTAATATTTCTGAAGAAGATATTTTGAATAAACGATGTACAAATTAACACTATTA 286  
QY 320 gggatcgcttgatgtttctgtacttttagggagcctcttaagttacattagggaaact 379  
DB 287 ATTATCCCTAGCTAGCAGATTTTGAGATTAGTAAATTAATACTATTAATAAATAAATG 346  
QY 380 ctacagcggtcaactcgttgggtttatccggaggttaaggtactactgtaaatgcaaatg 439  
DB 347 ATGTTTATTTTGTAAATGATAGTGTTCCTCAATAATACTACATAAATAAAGTGTATCC 406  
QY 440 aactaccacaaacgtttcttaagtaa 464  
DB 407 ATATCACAAATGTGACTATATAATAA 431

RESULT 15  
US-09-092-409-57  
; Sequence 57, Application US/09092409  
; Patent No. 6159478  
; GENERAL INFORMATION:



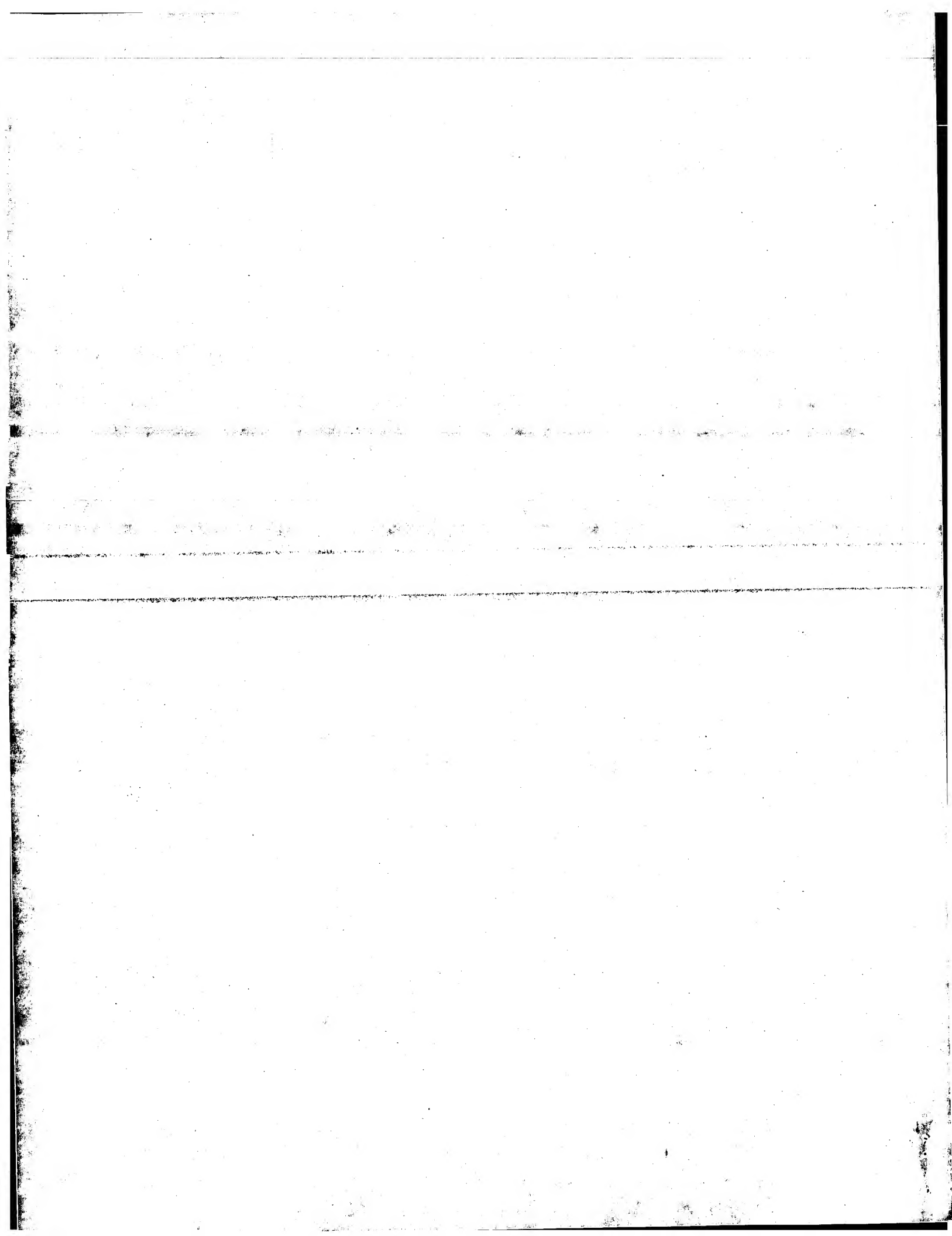
Fri Feb 8 08:44:54 2002

us-09-391-606-14.rni

Page 9

Search completed: February 7, 2002, 15:54:46  
Job time: 400 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 21:32:32 ; Search time 7419.31 Seconds  
(without alignments)  
1594.636 Million cell updates/sec

Title: US-09-391-606-14

Perfect score: 1101

Sequence: 1 atgttgctgtagggaaccc.....tatctggtcagttcagattc 1101

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estfun:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estom:\*
- 5: em\_estpl:\*
- 6: em\_estba:\*
- 7: em\_estro:\*
- 8: em\_estov:\*
- 9: em\_htc:\*
- 10: gb\_est1:\*
- 11: gb\_est2:\*
- 12: gb\_htc:\*
- 13: gb\_gss:\*
- 14: em\_gss\_fun:\*
- 15: em\_gss\_hum:\*
- 16: em\_gss\_inv:\*
- 17: em\_gss\_pln:\*
- 18: em\_gss\_pro:\*
- 19: em\_gss\_rtd:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	42.2	3.8	566	13	AQ778622 HS_2235_A
2	39.8	3.6	1101	13	CNS0039G
3	39.2	3.6	675	13	AZ840016
C 4	39	3.5	690	10	AU034107
5	38.4	3.5	1321	11	BG490965
C 6	38.2	3.5	296	10	AA381424
7	38.2	3.5	517	10	AW675282
8	38.2	3.5	534	10	AA431622
9	38	3.5	593	11	BF263367
10	37.6	3.4	869	13	BH131620
11	37	3.4	682	13	AQ925118
12	37	3.4	877	13	AZ671757

13	35.4	3.3	584	10	BE544037
14	36.2	3.3	539	13	AQ818532
15	36.2	3.3	1015	13	AZ686763
16	36.2	3.3	1041	13	CNS078LL
C 17	36	3.3	1101	13	CNS00LOO
C 18	35.8	3.3	418	11	BG153882
C 19	35.8	3.3	489	11	BI323992
C 20	35.8	3.3	749	13	AQ935992
C 21	35.8	3.3	852	13	CNS06229
C 22	35.8	3.3	887	13	AZ538399
23	35.4	3.2	274	10	BB424700
C 24	35.4	3.2	338	11	D71816
C 25	35.4	3.2	402	13	AQ477797
C 26	35.4	3.2	529	10	AW291247
C 27	35.2	3.2	254	10	BB081082
28	35.2	3.2	897	13	AZ677008
29	35.2	3.2	993	10	AL528462
C 30	35.2	3.2	994	13	AZ683097
C 31	35.2	3.2	1039	13	CNS07CAV
C 32	35.2	3.2	1101	13	CNS00FDD
C 33	35	3.2	310	10	AV277883
34	35	3.2	504	13	AZ617558
35	35	3.2	544	10	AW745780
36	35	3.2	771	11	BG584308
37	34.8	3.2	386	13	AQ442864
38	34.8	3.2	617	11	BE918310
39	34.8	3.2	700	10	AL503431
40	34.8	3.2	700	10	AL505428
41	34.8	3.2	786	11	BG218690
42	34.8	3.2	803	11	BG417399
43	34.8	3.2	985	10	BE412873
44	34.8	3.2	1041	13	CNS07BMO
45	34.6	3.1	371	13	AQ134473

## ALIGNMENTS

RESULT 1  
 AQ778622/C  
 LOCUS: 566 bp DNA GSS 02-AUG-1999  
 DEFINITION: HS\_2235\_A2\_B08\_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2235 Col=16 Row=C, DNA sequence.  
 ACCESSION: AQ778622  
 VERSION: AQ778622.1 GI:5681582  
 KEYWORDS: GSS.  
 SOURCE: human.  
 ORGANISM: Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE: 1 (bases 1 to 566)  
 AUTHORS: Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Koller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.  
 TITLE: Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 JOURNAL: Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 MEDLINE: 99380589  
 COMMENT: Contact: Mahairas GG, Wallace JC, Hood L  
 High throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones may be purchased from Research Genetics (info@resgen.com).  
 BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 2235 row: C column: 16  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 566.  
 Location/Qualifiers  
 1. .566







M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D. Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0136 row: D column: 20

Seq primer: CCTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 675.

Location/Qualifiers

1. .675

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="UUCG2M0136D20"

/clone\_lib="Mouse 10kb plasmid UUCG1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42av; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (g14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adaptor mouse DNA was annealed to

adaptor vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 201 a 164 c 146 g 164 t

ORIGIN

Query Match 3.6%; Score 39.2; DB 13; Length 675;  
Best Local Similarity 51.1%; Pred. No. 2.2;  
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 164 tagatgcacctaataattttttatggggagccagcctactggtatcgctgctgcaaac 223

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 423 TAGAGGACACTGAACAGTACTCTGTCATTACCAACCACTTACTGCTTGGCATATAAGT 482

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 224 atactactgcgttagatagactaacccgcgcctacataagcattttacacgagagt 283

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 483 AAACCAACAGAGAGCACTGCCAGACAAATCCAAATGAGATCTCCGACATTCACATA 542

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 284 ggttcactaatgcaggcttcattgccttaaacatttgggagcgtttgttttctgta 343

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 543 TGTGAACCTAAGTAAGGCTTAGTGTCTTAACCACTGGGACTGGAACAATGCTTTAAGCA 602

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4

AU034107/c

LOCUS

DEFINITION

ACCESSION

AU034107 690 bp mRNA EST 28-APR-1999  
AU034107 Dictyostellium discoideum SL (H.Urushihara) Dictyostellium  
discoideum cDNA clone SLB750, mRNA sequence.  
AU034107

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan

Email: d402hu@sakura.cc.tsukuba.ac.jp

PROJECT = 'Dictyostellium discoideum CDNA project in Japan'

POLYA-No.

Location/Qualifiers

1. .690

/organism="Dictyostellium discoideum"

/strain="AX4"

/db\_xref="taxon:44689"

/clone="SLB750"

/clone\_lib="Dictyostellium discoideum SL (H.Urushihara)"

/dev\_stage="slug"

BASE COUNT 225 a 141 c 148 g 176 t

ORIGIN

Query Match 3.5%; Score 39; DB 10; Length 690;

Best Local Similarity 46.8%; Pred. No. 2.6;

Matches 123; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 567 atatgcacgtcttaaacctaaagttaagaacttaagtgtatctgttaacgtatcgcaatt 626

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 413 ACAAGCAGAGACCAATTGAGAAGTTCAGCAGTTACAGCCATCGAATGCAATGAGACC 354

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 627 ctctgttaacaaacccagggcgtataaaggcgttcttcccttgcacacagacgtgg 686

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 353 TTCTTCTGTACACCATGGGATCTCTTGACAGATGCTGCTTGTGGCATTATCTACCAA 294

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 687 cgtagcaacagctactggaacaaagctcgcgaccatcaattatcatgaatggaatgag 746

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 293 ACCTTTAGTGGCTTGGGCACACAGATTTTGCACGACAGATTAACCTTCAGCCCAAGTGG 234

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 747 agcctctctatcttacagactaaactcttttagtgcatacatattggagatggtctcg 806

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 233 ATCTTTCTTATAGAGTGGACCAACCATCACTTCTTCTACCTTTTCGACACGTTCTCTTG 174

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 807 agcaacttttgatgctgataaca 829

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 173 TGTAAAGTGGCGGCACCAACCA 151

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Human

EST

602519759F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4638326 5',

mRNA sequence.

BG490965

BG490965

EST...

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1321)

AUTHORS

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

EST. 27-MAR-2001  
Dictyostellium discoideum  
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.  
1 (bases 1 to 690)  
Moriol, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,  
Yoshino, R., Miura, B., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,  
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.  
The Dictyostellium developmental cDNA project: generation and  
analysis of expressed sequence tags from the first-finger stage of  
development  
DNA Res. 5 (6), 335-340 (1998)  
99156227  
Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
Email: d402hu@sakura.cc.tsukuba.ac.jp  
PROJECT = 'Dictyostellium discoideum CDNA project in Japan'











```

1. 009
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica sheared DNA"
/notes="Vector: pHO51; Site1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica"

```

Seq primer: T7  
Class: BAC ends.

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-302M22"
/clone_lib="RPCI-23"

```







BASE COUNT	pBACø3.6 vector at ECORI sites"				10 others
	139 a	115 c	131 g	144 t	
ORIGIN					
	Query Match	3.3%;	Score 36.2;	DB 13;	Length 539;
	Best Local Similarity	49.2%;	Pred. No. 15;		
	Matches	95;	Conservative	0;	Mismatches 98; Indels 0; Gaps 0;
QY	857	cagctgttttaacttaactgc	atggaaacctcttttactag	aaaaatgccacagcattgt	916
Db	116	CAGTTGCGTGACTCTATT	TGGCAGGACATAAAGCC	TGGTCAATCTTTCAGGGTCAA	175
QY	917	ctactactgatttggtctcag	acttcaatgcaattgtttct	gtcagatcaacaagttta	976
Db	176	AAGAGACTGATATTTGGT	TTCATGCTCATGTCATG	ATGATGACGCCCTCACATCT	235
QY	977	aatctagaaaaagcttgttg	gagttactctaggagctact	ttagttgatgctgataaaat	1036
Db	236	AATCTAAGATTCTCGTGC	CTTCTCTAATGCCCAATG	GGTTTCATCTTGCCACATGCC	295
QY	1037	caacttactgcaga	1049		
Db	296	CAGATAGAGCTGA	308		

```

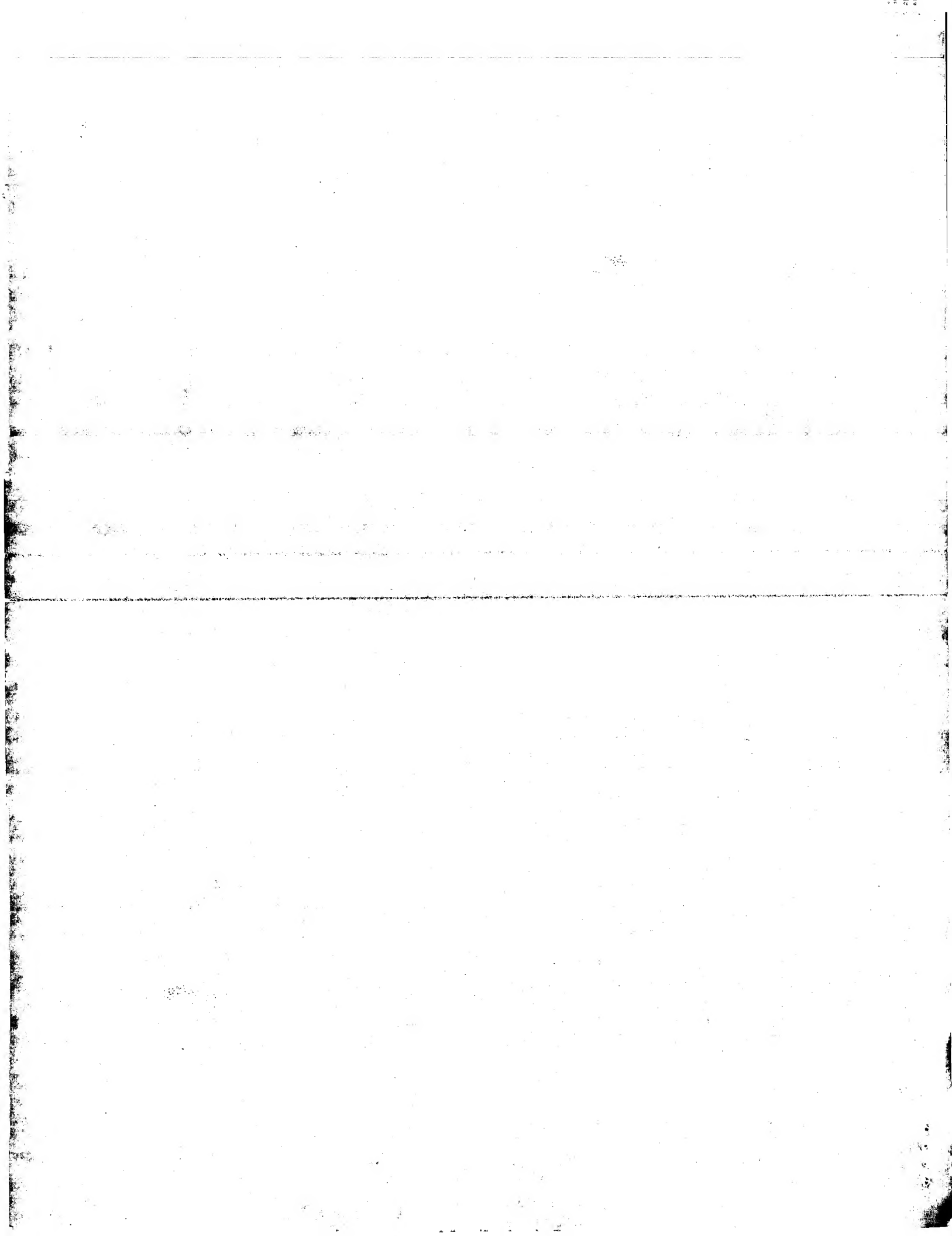
RESULT 15
AZ686763 1015 bp DNA GSS 14-DEC-2000
LOCUS ENT1568TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION AZ686763.1 GI:11823909
VERSION AZ686763
KEYWORDS Entamoeba histolytica.
SOURCE Entamoeba histolytica
ORGANISM Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1. (bases 1 to 1015)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: enta@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 86
High quality sequence stop: 556.
Location/Qualifiers
1. 1015
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a
method described by Clark and Diamond (Clark, A. C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.

```











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 20:23:51 ; Search time 172 Seconds  
(without alignments)  
184.541 Million cell updates/sec

Title: US-09-391-606-7  
Perfect score: 1166  
Sequence: 1 MTKKHVAVVVEGILNRLPKQ.....DKLGSDFTFRKFDLGIISAF 217

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Maximum Match 0%  
Post-processing: Minimum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL.17.\*
- 1: sp.archaea.\*
  - 2: sp.bacteria.\*
  - 3: sp.fungi.\*
  - 4: sp.human.\*
  - 5: sp.invertebrate.\*
  - 6: sp.mammal.\*
  - 7: sp.mhc.\*
  - 8: sp.organelle.\*
  - 9: sp.phage.\*
  - 10: sp.plant.\*
  - 11: sp.podent.\*
  - 12: sp.virus.\*
  - 13: sp.vertibrate.\*
  - 14: sp.unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1162	99.7	422	2 Q9Z7H6	Q9Z7H6 chlamydia p
2	1162	99.7	422	2 Q9J5B3	Q9J5B3 chlamydia p
3	1156	99.1	715	2 Q46166	Q46166 chlamydia p
4	778	66.7	432	2 Q84628	Q84628 chlamydia t
5	667	57.2	207	2 Q46380	Q46380 chlamydia t
6	92.5	7.9	901	12 Q9Q3F9	Q9Q3F9 bovine aden
7	88.5	7.6	830	12 Q84501	Q84501 paramecium
8	88.5	7.6	902	12 Q91ZAT	Q91ZAT bovine aden
9	85.5	7.3	684	2 Q9L1F8	Q9L1F8 streptomyce
10	85.5	7.3	836	12 Q9WAX4	Q9WAX4 chloroflex v
11	84	7.2	453	5 Q44021	Q44021 plasmodium
12	83.5	7.2	246	2 Q9L121	Q9L121 streptomyce
13	83	7.1	349	12 Q87696	Q87696 chimpanzee
14	83	7.1	743	1 Q58900	Q58900 pyrococcus
15	83	7.1	2233	2 Q9KG00	Q9KG00 streptococc
16	82.5	7.1	568	2 Q915N0	Q915N0 pseudomonas
17	82	7.0	1176	12 Q98587	Q98587 paramecium
18	81.5	7.0	395	5 Q26913	Q26913 trypanosoma
19	81.5	7.0	449	5 Q00819	Q00819 trypanosoma

20	81.5	7.0	900	12 Q91ZAT	Q91ZAT bovine aden
21	81	6.9	1128	2 Q9FDC0	Q9FDC0 bacillus th
22	80.5	6.9	214	2 P72800	P72800 synchocyst
23	80.5	6.9	409	2 O66864	O66864 aquifex aeo
24	80.5	6.9	447	2 Q9AGU3	Q9AGU3 pseudomonas
25	80.5	6.9	469	2 O50499	O50499 streptomyce
26	80.5	6.9	902	12 Q9Q3G0	Q9Q3G0 bovine aden
27	80	6.9	644	1 Q9V057	Q9V057 pyrococcus
28	79.5	6.8	290	2 Q59078	Q59078 alteromonas
29	79.5	6.8	910	12 Q711A9	Q711A9 bovine aden
30	79	6.8	195	2 Q9KLF5	Q9KLF5 vibrio chol
31	79	6.8	473	2 Q9RB20	Q9RB20 erwinia car
32	79	6.8	482	10 Q80804	Q80804 arabidopsis
33	79	6.8	499	2 Q9ZL55	Q9ZL55 helicobacte
34	79	6.8	648	5 Q9VKU7	Q9VKU7 drosophila
35	79	6.8	650	5 Q9ULJ7	Q9ULJ7 drosophila
36	78.5	6.7	387	5 Q9VKT8	Q9VKT8 drosophila
37	78.5	6.7	412	2 Q50371	Q50371 methylophil
38	78.5	6.7	658	10 Q9FTM2	Q9FTM2 oryza sativ
39	78.5	6.7	3796	6 Q97699	Q97699 bos taurus
40	78.5	6.7	3796	6 Q9TTK4	Q9TTK4 bos taurus
41	78	6.7	753	10 Q9SUA4	Q9SUA4 arabidopsis
42	78	6.7	925	1 Q9HLE8	Q9HLE8 thermoplasma
43	77.5	6.6	696	8 Q9TIT2	Q9TIT2 phacelia ra
44	77.5	6.6	3186	4 Q9H133	Q9H133 homo sapien
45	77	6.6	505	1 Q28945	Q28945 archaeoglob

## ALIGNMENTS

RESULT 1

Q9Z7H6 PRELIMINARY; PRT; 422 AA.

AC Q9Z7H6

DT 01-MAY-1999 (TRENBLrel. 10, Created)

DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DE CHLPN 76 KDA HOMOLOG\_2 (CT623).

GN CPN0729 OR CP0017.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

QX NCBI\_TaxID=83558;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029;

RX MEDLINE=99206606; PubMed=10192388;

RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,

RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";

RL Nat. Genet. 21:385-389(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AR39;

RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,

RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,

RA Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,

RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;

RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).

DR EMBL; AE001654; AAD18868.1; -

DR EMBL; AE002165; AAF37913.1; -

DR TIGR; CP0017; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 422 AA; 48374 MW; ICFESB64B0ED3084 CRC64;

Query Match 99.7%; Score 1162; DB 2; Length 422;  
Best Local Similarity 99.5%; Pred. No. 6.7e-99;  
Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia. NCBI\_TaxID=83558;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=94156481; PubMed=7509320;  
 RX Perez-Melgosa M., Kuo C.C., Campbell L.;  
 RA "Isolation and characterization of a gene encoding a Chlamydia  
 RT pneumoniae 76-kilodalton protein containing a species-specific  
 RT epitope.";  
 RL Infect. Immun. 62:880-886(1994).  
 DR EMBL; L23921; AAA32117.1;  
 SQ SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;

Query Match 99.1%; Score 1156; DB 2; Length 715;  
 Best Local Similarity 99.5%; Pred. No. 4.9e-98;  
 Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMKYKVCWQWLV 60  
 Db 206 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMKYKVCWQWLV 265

QY 61 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTLNKENLAWFIGTGLGRLKAGDWSA 120  
 Db 266 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTLNKENLAWFIGTGLGRLKAGDWSA 325

QY 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 180  
 Db 326 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 385

QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217  
 Db 386 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 422

RESULT 2  
 Q9JSB3 ID Q9JSB3 PRELIMINARY; PRT; 422 AA.  
 AC Q9JSB3;  
 DT 01-OCT-2000 (TremBLrel. 15, Created)  
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)  
 DE CHLPN 76 KDA HOMOLOG\_2 (Cf6233).  
 GN CPJ0729.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL; AF002547; BAA98936.1;  
 SQ SEQUENCE 422 AA; 48373 MW; FDC2F564B2679C0E CRC64;

Query Match 99.7%; Score 1162; DB 2; Length 422;  
 Best Local Similarity 99.5%; Pred. No. 6.7e-99;  
 Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMKYKVCWQWLV 60  
 Db 206 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMKYKVCWQWLV 265

QY 61 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTLNKENLAWFIGTGLGRLKAGDWSA 120  
 Db 266 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTLNKENLAWFIGTGLGRLKAGDWSA 325

QY 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 180  
 Db 326 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 385

QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217  
 Db 386 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 422

RESULT 3  
 Q46166 ID Q46166 PRELIMINARY; PRT; 715 AA.  
 AC Q46166;  
 DT 01-NOV-1996 (TremBLrel. 01, Created)  
 DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)  
 DE 76 KDA PROTEIN.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia. NCBI\_TaxID=83558;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=94156481; PubMed=7509320;  
 RX Perez-Melgosa M., Kuo C.C., Campbell L.;  
 RA "Isolation and characterization of a gene encoding a Chlamydia  
 RT pneumoniae 76-kilodalton protein containing a species-specific  
 RT epitope.";  
 RL Infect. Immun. 62:880-886(1994).  
 DR EMBL; L23921; AAA32117.1;  
 SQ SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;

Query Match 99.1%; Score 1156; DB 2; Length 715;  
 Best Local Similarity 99.5%; Pred. No. 4.9e-98;  
 Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMKYKVCWQWLV 60  
 Db 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMKYKVCWQWLV 60

QY 61 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTLNKENLAWFIGTGLGRLKAGDWSA 120  
 Db 61 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTLNKENLAWFIGTGLGRLKAGDWSA 120

QY 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 180  
 Db 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 180

QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 216  
 Db 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 216

RESULT 4  
 O84628 ID O84628 PRELIMINARY; PRT; 432 AA.  
 AC O84628;  
 DT 01-NOV-1998 (TremBLrel. 08, Created)  
 DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
 DE CHLPN 76 KDA HOMOLOG.  
 GN CT623.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=D/UW-3/CX;  
 RX MEDLINE=9900809; PubMed=9784136;  
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis.";  
 RL Science 282:754-759(1998).  
 DR EMBL; AE001333; AAC68227.2;  
 KW Complete proteome.  
 SQ SEQUENCE 432 AA; 48365 MW; 46B3BA0A193A86BA CRC64;

Query Match 66.7%; Score 778; DB 2; Length 432;  
 Best Local Similarity 63.8%; Pred. No. 1.5e-63;  
 Matches 148; Conservative 26; Mismatches 42; Indels 15; Gaps 4;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFV-----PSETSTTEKAATNAMKYKVCV 55  
 Db 202 MAEKEYAVWVEAILNKLPGNFVVKSVVDNLTAKTNDPADASAAQPAKPT-KYDYL 260

QY 56 WQWLGVKHSQVPWINGQKKPLLYLYGAFNLNPLAK-----ATKTT--LNGKENLAWFI 105  
 Db 261 WQWLGVKSTAMPWFNGQTKNLYGAYLFNPLAEIPENKQSTPTPTTKITNGKENHAWFI 320



QY 106 GCTGLGRRKAGDWSATVRYEYVWLVGKHSQVWPNWINGOKKPLYLYGAFMLNPLAKATKT 165  
 Db 321 GCSLGGVRRAGDWSATVRYEYVWLVGKHSQVWPNWINGOKKPLYLYGAFMLNPLAKATKT 380  
 QY 166 FTYNYGFSALYMGITDLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217  
 Db 381 FTYNYGFSYQFVGMGLTDSVFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 432

RESULT 5  
 Q46380 PRELIMINARY; PRT; 207 AA.  
 AC Q46380;  
 DT 01-NOV-1996 (TremBLrel. 01, Created)  
 DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)  
 DE 76 KDA PROTEIN HOMOLOG.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Myers G.S.A., Sriprakash K.S.;  
 RT "A membrane-associated gene cluster of Chlamydia trachomatis."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U50732; AAD08717.1;  
 SQ SEQUENCE 207 AA; 23254 MW; 9887DB846DCB55C8 CRC64;

Query Match 57.2%; Score 667; DB 2; Length 207;  
 Best Local Similarity 61.6%; Pred. No. 9.4e-54;  
 Matches 130; Conservative 24; Mismatches 37; Indels 20; Gaps 5;

QY 23 VKCSVVDNNTFV-----PSETSTTEKATNAMYKVCVWLVGKHSQVWPNWINGOKKPLY 77  
 Db 1 MKTSSVVDNNTLTAKTNDPADASAAQAPKNT-KYDYLWQVLGKSTAMPWFNGQTKNLY 59

QY 78 LYGAFMLNPLAK-----ATKTTLNKGNLAWFIGGTLGRLKAGDWSATVRYEY 126  
 Db 60 TYGAYLFNPLAEIPENWQKSTTPATKIT-NGKNHAWFICGSLGVRAGDWSATVRYEY 118

QY 127 VEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAIYMGITDLSLF 186  
 Db 119 VKLR--PEIDVAGIGRGNQMYWFAQAIAKGLDPRKESNGFTNYKGVSYQFVGMGLTDSVSF 176

QY 187 RAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217  
 Db 177 RAYGAYSKPANDKLGSDFTFRKFDLGIISAF 207

RESULT 6  
 Q903F9 PRELIMINARY; PRT; 901 AA.  
 AC Q903F9;  
 DT 01-MAY-2000 (TremBLrel. 13, Created)  
 DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)  
 DE HEXON (FRAGMENT).  
 OS bovine adenovirus 6.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Adenovirus.  
 OX NCBI\_TaxID=111167;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=671130;  
 RA Lehmkuhl H.D., Hobbs L.A.;  
 RT "Partial sequence of the bovine adenovirus type 6, hexon gene."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF207659; AAF20945.1;  
 DR HSSP; P03277; 1DHX.  
 DR InterPro; IPR000736; Adeno\_hexon.  
 DR Pfam; PF01065; Adeno\_hexon; 1.  
 DR ProDom; PD002815; Adeno\_hexon; 1.

FT NON\_TER 1 901  
 FT NON\_TER 901  
 SQ SEQUENCE 901 AA; 101998 MW; 5E74B05C8246F353 CRC64;

Query Match 7.9%; Score 92.5; DB 12; Length 901;  
 Best Local Similarity 21.3%; Pred. No. 5.2;  
 Matches 40; Conservative 28; Mismatches 77; Indels 43; Gaps 7;

QY 34 VPSETSTTEKATNAMYKVCVWLVGKHSQVWPNWINGOKKPLYLYGAFMLNPLAKATKT 93  
 Db 602 IPPNTNTVVVNIPIRDSWGAFCGWSFNRLKASETPMGATKDPNFLYSGSI--PLDGT-- 657

QY 94 TLNGKENLAWFIGGTLGRLKAGDWSATVRYEYVWLVGKHSQVWPNWINGOKKPLYLYGAFMLNPLAKATKT 142  
 Db 658 -----FYLSTHF--QVSIQNDSSVPWPGDRLIPNWFEEKRDNIDAEGYNMSQ 706

QY 143 GNLLKFWFAQAIAANYDPKEANSFTNYKGFSAIYMGITDLSL-----FRAY 189  
 Db 707 SNITRDWYMIQMAANYN---QAYQGYKLPSQHKYGFTEIENFQPMTRQVPMYGDGVYDLY 762

QY 190 GAYSKPAN 197  
 Db 763 NAYLSNPN 770

RESULT 7  
 Q84501 PRELIMINARY; PRT; 830 AA.  
 ID Q84501  
 AC Q84501; Q84502;  
 DT 01-NOV-1996 (TremBLrel. 01, Created)  
 DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
 DE PBCV-1 CHITINASE.  
 GN A181/182R.  
 OS Paramacium bursaria chlorella virus 1 (PBCV-1).  
 OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.  
 OX NCBI\_TaxID=10506;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96187795; PubMed=8614977;  
 RA Lu Z., Li Y., Que Q., Kutish G.F., Rock D.L., Van Etten J.L.;  
 RT "Analysis of 94 kb of the chlorella virus PBCV-1 330-kb genome: map  
 positions 88 to 182."  
 RL Virology 216:102-123(1996).  
 DR EMBL; U42580; AAC96549.2;  
 DR HSSP; P07986; 1EXG.  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 SQ SEQUENCE 830 AA; 90929 MW; 12D20FFFEDE008076 CRC64;

Query Match 7.6%; Score 88.5; DB 12; Length 830;  
 Best Local Similarity 22.4%; Pred. No. 11;  
 Matches 38; Conservative 19; Mismatches 44; Indels 69; Gaps 6;

QY 4 KHYAWVVEGILNR-----LPKQFVKCSVVDWNTFVPSSTT-----EKAATNAMY 51  
 Db 50 ENFTWFSEGDVLRKGNKVTMPK-----DWNMSIPAGTKIIPFGGVKALPGLNLY 100

QY 52 KYCVWQVLGKHSQVWPNWINGOKKPLYLYGAFMLNPLAKATKTTLNKENLAWFIGGTG 111  
 Db 101 NOIL--PLVGK-----DP-----S 112

QY 112 LKAGDWSATVRYEYVWLVGKHSQVWPNWINGOKKPLYLYGAFMLNPLAKATKTTLNKENLAWFIGGTG 161  
 Db 113 LAKRGKSSKAVAPYVDACAFPTDLPALSKAGLKFETLAFITADSNNK 162

RESULT 8  
 Q91ZA7 PRELIMINARY; PRT; 902 AA.  
 ID Q91ZA7  
 AC Q91ZA7;



DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE HEXON (FRAGMENT).  
OS bovine adenovirus 8.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Adenovirus.  
OX NCBI\_TaxID=120509;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MISK/67;  
RA Lemkuhl H.D., Hobbs L.A.;  
RT "Partial sequence of the bovine adenovirus type 8 hexon gene."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF238233; AAF63489.1; -  
DR InterPro; IPR001570; Peptidase\_M4.  
DR Pfam; PF01065; Adeno\_hexon; 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.  
DR NON\_TER 902 902  
FT NON\_TER 1  
SQ SEQUENCE 902 AA; 102443 MW; 450C5E09AD3A2243 CRC64;

Query Match 7.6%; Score 88.5; DB 12; Length 902;  
Best Local Similarity 20.7%; Pred. No. 12;  
Matches 39; Conservative 29; Mismatches 77; Indels 43; Gaps 7;  
QY 34 VPSETSTTEKATNAMKYCYWQWLVGKHSQVPWINGOKKPLYLYGAFLLNPLAKATKT 93  
DB 603 IPNFTSTVVNVPDRSWAFRCWSNVRKATETPMIGATKDPNFLYSGSI--PYLDGI-- 658  
QY 94 TLNGKENLAWFTGGLGRKAGDSATRYEYVEALSVP-----EIDVSG--IGR 142  
DB 659 -----FYLTHFF--QRVSIQWSSVPWPGDRLIPNWFIEIKRDYSVDAEGYNMSQ 707  
QY 143 GNLLKFWFAQAATAANDPKANSFTNYKGFSAIYMGITDLSL-----FRAY 189  
DB 708 CNITKDWYMIQMAYN-----QAYQGYKLPVSHKYGFLENFQPMRSQVPIYGNLDFOLY 763  
QY 190 GAYSPAN 197  
DB 764 SAYIKSPN 771

RESULT 9  
Q9L1F8 PRELIMINARY; PRT; 684 AA.  
AC Q9L1F8;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE PUTATIVE NEUTRAL ZINC METALLOPROTEASE.  
GN SC3D11.04C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Brown S.P., Harris D.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rastam M.A.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL157916; CAB76001.1; -  
DR InterPro; IPR001570; Peptidase\_M4.  
DR InterPro; IPR000130; Zn\_MTPeptidase.  
DR Pfam; PF01447; Peptidase\_M4; 1.  
DR PRINTS; PR00730; THERMOLYSIN.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Protease; Metalloprotease.  
SQ SEQUENCE 684 AA; 71289 MW; 594F70DDAB732BCF CRC64;

Query Match 7.3%; Score 85.5; DB 2; Length 684;  
Best Local Similarity 23.8%; Pred. No. 16;  
Matches 43; Conservative 23; Mismatches 62; Indels 53; Gaps 10;  
QY 55 VWQWLVGKHSQVPWINGOKKPL-YLYGAFLLNPLAKATKTTLNGKENLAWFTGGLG-- 111  
DB 417 VGDYLGEEID---INGDGTPLRYM-----DKPSKDGASKDSWYSG--IGVD 459  
QY 112 -----LRKAGDSATRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANY 158  
DB 460 VHYSSGPANHFYLLSESGSGTKTINGVYNSPTSDGLPVTGIGRAKAKEIWF-RALTTFK 518  
QY 159 DPKEANSFTNYK-----FSALYMGITDLSLSPRA-----YCAYSKPANDKLGSDFT 205  
DB 519 -----TSTNTYAGARTGTLLAAAGELYG-TDSAETAVANAWAGINVGARPGGDPDGT 572  
QY 206 F 206  
DB 573 F 573

RESULT 10  
Q9WAX4 PRELIMINARY; PRT; 836 AA.  
ID Q9WAX4  
AC Q9WAX4; (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE VCHTI-1.  
GN VCHTI-1.  
OS Chlorella virus.  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.  
OX NCBI\_TaxID=10507;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CVK2;  
RA Hiramatsu S., Yamada T.;  
RT "Chlorella virus vChTI-1 gene, complete cds."  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB022343; BAA78554.1; -  
DR HSP; P07986; 1EXG.  
DR InterPro; IPR002965; P-rich\_extensn.  
DR PRINTS; PR01217; PRICHEXTENS.  
SQ SEQUENCE 836 AA; 91356 MW; 6C283A494BA908E2 CRC64;

Query Match 7.3%; Score 85.5; DB 12; Length 836;  
Best Local Similarity 21.8%; Pred. No. 21;  
Matches 37; Conservative 20; Mismatches 44; Indels 69; Gaps 6;  
QY 4 KHYAVVVEGILNR-----LPKQFFVKCSVVVDWNTFVPSFTT-----EKAATNAMKY 51  
DB 50 ENFTWFSEGDLYRKGNKVTMPK-----DWNMSIPAGTTKIIPFGGVKALPGNLKY 100  
QY 52 KYCVWQWLVGKHSQVPWINGOKKPLYLYGAFLLNPLAKATKTTLNGKENLAWFTGGLG 111  
DB 101 NQIL--PLVGK-----DP-----S 112  
QY 112 LRKAGDSATRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPK 161  
DB 113 LAKRGKWTSKAVAPYVDACAFPTPDLPAISKASGLKFFTLATITADSNK 162



Query Match 7.1%; Score 83; DB 12; Length 349;  
 Best Local Similarity 20.9%; Pred. No. 11;  
 Matches 46; Conservative 26; Mismatches 82; Indels 66; Gaps

```

[2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);

```



QY 4 KHYAVVEGILNRL---PKOFFVKCSVDWNTFVPSSTTEKAATNAMKYKCVWQWLV 60  
 DB 124 KHY-W--DAIRFYCAPPGVALLKNDNTNSGFWPNCSSKVVASCTRTMETQTPWFGFN 180  
 QY 61 GKHSQ---VPWINGQKPLV---LYGAFLNPLAKATKTTL-----NGK 98  
 DB 181 GTRARENTHYWHGRDNRPIFSLNQYVNLNMCRRPCNKVPLVPIISGWVHSLPINDS 240  
 QY 99 ENLAWFTGGTGLGURKAGW-----SATVRY-EVEALSVPEDVSGIGRG----- 143  
 DB 241 PNQAWCWFG-----GKWKDAIKEYKQALVHPRTYGANNITDEINLTPPGGDPVTF 292  
 QY 144 -----NLLKFWFAQAIAAAYDPKEANSFTNY 169  
 DB 293 MWTNCRGEFLYCKMWNFLNWDRTANQKPKHHK-RNY 331

RESULT 14  
 ID O58900 PRELIMINARY; PRT; 743 AA.  
 AC O58900;  
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
 DE 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
 DE HYPOTHETICAL 81.5 KDA PROTEIN PH1196.  
 GN PH1196.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OX NCBI\_TaxID=53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE=98344137; PubMed=9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Masuchi Y., Shizuya H., Kikuchi H.,  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3."  
 RL DNA Res. 5:55-76(1998).  
 KW EMBL; AP000005; BAA30296.1;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 743 AA; 81475 MW; 6FEAF83BC01F3673 CRC64;

Query Match 7.1%; Score 83; DB 1; Length 743;  
 Best Local Similarity 18.5%; Pred. No. 30;  
 Matches 42; Conservative 35; Mismatches 76; Indels 74; Gaps 9;

QY 9 VVEGILNRLPKOFFVKCSVDWNTFVPSSTTEKAATNAMKYKCVWQWLVGKHSQVWP 68  
 DB 314 IIEELFNAPKOFIEVAEVSRYREI-----YYI-----NNGQW 346  
 QY 69 INQKPLLYLYGAFLNPLAKATKTTLNGKENLAWFTGGTGLGURKAGDWSATVRYEYVE 128  
 DB 347 IDHKKEP-----SAKYITLSGYRHIG-----GGVVVIGGFGAGSLPADYKD 387  
 QY 129 A-----LSVPEIDVSGIGRGNLLKFW-----FAQAIAAYDPKEANSFTN 168  
 DB 388 APEAAKFTAFQTKVSGASSNLRLWLSAPLHMLNATVAYRFYVQANPGYDFDPSPSIIDE 447  
 QY 169 YKGFSAIYMGITDLSLFRAYGAYSKPAN--DKLGSDFTFKFDLGI 213  
 DB 448 ----ARIVYRITDKGTWEL--AYALPSGVNLDGLTTGKRLDASV 488

RESULT 15  
 Q9KGUO  
 ID Q9KGUO PRELIMINARY; PRT; 2233 AA.  
 AC Q9KGUO;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE BETA-GALACTOSIDASE PRECURSOR.  
 GN BGAA.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RN7/87;  
 RA Zaehner D., Hakenbeck R.;  
 RT "The Streptococcus pneumoniae beta-galactosidase is a surface  
 protein."  
 RL J. Bacteriol. 0:0-0(2000).  
 DR EMBL; AF282987; AAF97242.1;  
 DR InterPro; IPR001649; Glyco\_hydro\_2.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PRINTS; PR00132; GLHYDRLASE2.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 SQ SEQUENCE 2233 AA; 247290 MW; 8170D06584D0A137 CRC64;

Query Match 7.1%; Score 83; DB 2; Length 2233;  
 Best Local Similarity 23.4%; Pred. No. 1.3e+02;  
 Matches 57; Conservative 28; Mismatches 81; Indels 78; Gaps 14;

QY 32 TFVPSSTTEKAATNAMKYKCVWQWLVGKHSQVWPWINGOKKPLLYLYGAF-----MNPL 87  
 DB 483 THNPASEQTQIAAELGL-----LVOEAFDTWYGG-KRP-YDYGRRFEKDATHP 531  
 QY 88 AKA-----TKTTL-NGKENLA---WFIGTGLGURKAGDWSATVRYEYVEALSVP 135  
 DB 532 ARKGEKSDPDLRTMVERGKNNPAIFMWSIGNEIGEANGDAHSLATVK-RLVKV--IKDV 588  
 QY 136 DVSGIGRGNLLKFWFA-----QAIAAYDPKEANSFTNYKGFSAIY-----MY 178  
 DB 589 DKTRYVTMGADKFRFGNGSGGHEKIADELDAVGFN-----SEDNYKALRAKHPK 642  
 QY 179 GITUSLSFRAYGAYSKP-----ANDKLG-----SDTFKFDLGI 213  
 DB 643 GSETSSATRTGRSYRPERELKHSNGPERNEVEQSDYGNDRVWGKGTATATSWTFDRNAGY 702  
 QY 214 ISAF 217  
 DB 703 AGQF 706

Search completed: February 7, 2002, 21:41:09  
 Job time: 4638 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:30:32 ; Search time 12230.8 Seconds  
(without alignments)  
1873.512 Million cell updates/sec

Title: US-09-391-606-4  
Perfect score: 1389  
Sequence: 1 atggttaactctattgtcc.....aagttctatggtccgagct 1389

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_on.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1379.4	99.3	2238	6	AX045137	AX045137 Sequence
2	1369.4	98.6	2148	1	CHT76KDA	L3921 Chlamydia p
3	1359.4	97.9	2156	1	AX045131	AX045131 Sequence
c 4	1359.4	97.9	10421	1	AE001654	AE001654 Chlamydia
5	1359.4	97.9	12548	1	AE002165	AE002165 Chlamydia
c 6	1359.4	97.9	300550	1	AP002547	AP002547 Chlamydia
7	1347	97.0	1456	6	AX045135	AX045135 Sequence
8	1155.4	83.2	1852	6	AX045133	AX045133 Sequence
9	146.8	10.6	150	6	ARI44067	ARI44067 Sequence
c 10	56.6	4.1	10954	1	AE001333	AE001333 Chlamydia
c 11	54.4	3.9	7218	6	I66494	I66494 Sequence 14
12	53.6	3.9	1537	6	AX155814	AX155814 Sequence
13	51	3.7	5098	1	MCU51235	U51235 Mycoplasma
14	50.6	3.6	2085	6	AR008322	AR008322 Sequence
15	50.6	3.6	2085	6	AR027306	AR027306 Sequence
16	50.6	3.6	2085	6	AR035771	AR035771 Sequence
17	50.6	3.6	2085	6	AR037631	AR037631 Sequence
18	50.6	3.6	2085	6	AR078816	AR078816 Sequence
19	50.6	3.6	2085	6	AR079063	AR079063 Sequence
20	50.6	3.6	2085	6	AR084296	AR084296 Sequence
21	50.6	3.6	2085	6	I16740	I16740 Sequence 1
22	50.6	3.6	2085	6	I70472	I70472 Sequence 1
23	50.6	3.6	2085	6	I92332	I92332 Sequence 1
24	50.6	3.6	2086	1	STRSURPROA	M74122 Streptococc
c 25	50.2	3.6	14698	1	AE002357	AE002357 Chlamydia
26	49	3.5	423	1	MC049	Z33043 M. capricolu
27	48.8	3.5	10607	1	AE006310	AE006310 Lactococc
28	45.8	3.3	719	1	AF244521	AF244521 Streptoco
29	45.8	3.3	151847	2	AC022792	AC022792 Homo sapi
c 30	45.6	3.3	3720	14	S76368	S76368 ORF 5' of E
c 31	45.6	3.3	43658	14	HSV3PRGEN	H86409 Herpesvirus
c 32	45.6	3.3	112930	14	HSGEND	X64346 Herpesvirus
33	44.8	3.2	110000	2	PFMAL13P2_0	AL049185 Plasmodi
c 34	44.8	3.2	168310	2	AC090810	AC090810 Homo sapi
c 35	44.6	3.2	3882	10	AF104261	AF104261 Mus muscu
36	44.6	3.2	177205	2	AC079217	AC079217 Mus muscu
37	44.6	3.2	178419	2	AC074040	AC074040 Mus muscu
38	44.4	3.2	633	9	HSU23862	U23862 Human clone
39	44.4	3.2	149592	2	AC073077	AC073077 Homo sapi
40	44.4	3.2	173184	2	AC013583	AC013583 Homo sapi
41	44.2	3.2	175895	9	AC064862	AC064862 Homo sapi
42	43.8	3.2	463	3	PFA332C	M69163 P. falciparu
43	43.8	3.2	16377	3	AF202180	AF202180 Plasmodi
c 44	43.8	3.2	78578	2	PFMAL13P9	AL096783 Plasmodi
45	43.6	3.1	6744	1	SSEPFAA	X71880 S.suis (189

## ALIGNMENTS

RESULT 1

AX045137  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
Source

AX045137  
Sequence 7 from Patent WO0066739.  
AX045137  
AX045137.1 GI:11343736  
Chlamydia pneumoniae.  
Chlamydia pneumoniae  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
1 (bases 1 to 2238)  
Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.  
i(chlamydia) antigens and corresponding dna fragments and uses thereof  
Patent: WO 0066739-A 7 09-NOV-2000;  
Aventis Pasteur Limited (CA)  
Location/Qualifiers  
1..2238  
/organism="Chlamydia pneumoniae"  
/db\_xref="taxon:83558"

PAT

24-NOV-2000



CDS	766...2238	/note="unnamed protein product"	
		/codon_start=1	
		/transl_table=11	
		/db_xref="GI:11343737"	
		/protein_id="CAC17330.1"	
		/translation="MVNPIGPGPIDETERTPPADLSAQGLEASAAANKSAFAQRIAGAE	
		AKPKESKTSVERWSILRSVNALMSLADLGIASSNSSSTSRSDVDTTATPTP	
		PPPFIDYKQAOATPYDTIFTSLSADIQAALVSLQDAVTNKIDTATDETAIAAEW	
		ETKNADAVKVCQATTELAKYASDNQAILDSGLKLTSPDLLQALLQSVANNKAAELL	
		KEMQNDVPVFKTPAIQAQSLVDQDATATQEKDGNRAIRDYFAGQNASGAVENAKSN	
		NSISNIDSAKAATAATKQIAEAQKFPDSPILQEAQVMVIOAKDLKNIKPDGSDV	
		PNPQTTVGGSKQSSSIRSVMLDDAENETSLMSGFRMIMHFNFDSDQAA	
		QOELAAQARAAGKAGDDSAALALADAKALEAALGKAGQOQILNALGQIASAAVSA	
		GVLPLQQLVLRARYQAYVEQKLISEEDLSAVDHHHHH"	
BASE COUNT	689 a 459 c 497 g 593 t		
ORIGIN			
Query Match 99.3%; Score 1379.4; DB 6; Length 2238;			
Best Local Similarity 99.6%; Pred. No. 0;			
Matches 1383; Conservative 0; Mismatches 6; Indels 0; Gaps 0;			
QY	1	atggttaatcctattggtccaggctctatagacgaacacagacacacccctcccgagat	60
Db	766	ATGGTTAATCCTATTGGTCCAGGTCCTATAGACGAACACAGACGACACCTCCCGCAGAT	825
QY	61	ctttctgctcaaggattgagcagcagtcagcaacaataagagtcggaagctcaagaata	120
Db	826	CTTTCTGCTCAAGGATTGAGCGGAGTGCAGCAATAAGAGTGGCGAAGCTCAAGAATA	885
QY	121	gcaggtgcggaagctaaagcctaaagaatctaaagaccgattctgtagacgcatgagcgc	180
Db	886	GCAGGTGCGGAAGCTAAGCCTAAACAATCTAAGACCGATTCTGTAGACGATGAGCATC	945
QY	181	tgcggtctgcagtgaaatgctctatagtcgtgagcagataagctgggtattgcttctagt	240
Db	946	TTGCGTTTGCAGTGAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT	1005
QY	241	aacagctcgtctctactagcagatctgcagcgtggactcaacacagcagccagcct	300
Db	1006	ACAGCTGCTCTCTACTAGCAGATCTGCAGACGTGGACTCAACGACAGCGACCGCAGCT	1065
QY	301	acgcctctcccccacgctctgatgattataagactcaagcgcaaacagcgttaacgatact	360
Db	1066	ACGCTCTCCACCACCGCTTTCATGATTATAAGACTCAAGCGCAACAGCTTACGATACT	1125
QY	361	atctttacctaacaatcactagctgacatacaggtgtgttgtagcctccaggatgct	420
Db	1126	ATCTTTACCTCAACTACACTAGCTGACATACAGGCTGCTTTTGGTGAGCCTCCAGGATGCT	1185
QY	421	gtcactaataaagatacagcggctactgtaggaaacgcgaatcgctgcggagtgg	480
Db	1186	GTCACTAATAAAGATACACGGGCTACTGATGAGGAACACCGCAATCGCTCGGAGTGG	1245
QY	481	gaaactaagaatgccgatgcaattaaagtggcgcgcaaaattacagaattagcgaatat	540
Db	1246	GAAACTAAGAATGCCGATGCAGTTAAAGTTGCGCGCAAAATACAGAATTAGCGAAATAT	1305
QY	541	gcttcggataacacagcagcttctgactcttttagttaactgaactccttcgaaccttta	600
Db	1306	GCTTCGGATAACCAACGCAATCTTACTCTTTAGTAACTGACTTCCTTCGACCTCTTA	1365
QY	601	cgagctgctctctccaatctgtagcaacaatacaaacagcagctgagcttcttaagag	660
Db	1366	CAGGCTGCTCTTCTCCATCTCTGACCAACAATATACAAAGCAGCTGAGCTCTTAAGAG	1425
QY	661	atgcaagataaaccagtagtccaggggaaacgcctgcaattgctcaatcttttagttgat	720
Db	1426	ATGCAAGATAACCCAGTAGTCCCGGAAACCGCTGCAATTCCTCAATCTTTAGTTGAT	1485
QY	721	cgacagatgcttacagcgacacagatagagaaagatggaatgcgattgggagatcatat	780

RESULT 2

CHT76KDA	2148 bp	DNA	BCT	17-MAR-1994
LOCUS	Chlamydia pneumoniae	76 kda protein gene, complete cds.		
DEFINITION	Chlamydia pneumoniae			
ACCESSION	L23921			
VERSION	L23921.1	GI:435961		
KEYWORDS	76 kda protein.			
SOURCE	Chlamydia pneumoniae DNA.			
ORGANISM	Chlamydia pneumoniae			
REFERENCE	1 (bases 1 to 2148)			
AUTHORS	Perez-Melgosa, M., Kuo, C.-C. and Campbell, L.			
TITLE	Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76-kilodalton protein containing a species-specific epitope			
JOURNAL	Infect. Immun.	62, 880-886	(1994)	
MEDLINE	94156481			
FEATURES	Location/Qualifiers			
source	1..2148	/organism="Chlamydia pneumoniae"		
		/db_xref="taxon:83538"		
CDS	1..2148	/codon_start=1		
		/transl_table=11		
		/product="76 kDa protein"		
		/protein_id="AAA23117.1"		



```

/db_xref="GI:435962"
/translation="MTKKHYAWVVEGILNRLPKQFFVKVSVVDWNTFVPSESTTEKLA
ATNAMYKYCVQMQLVQKSPWINGKQKPLYLYGAFLMNP LAKATKTKTLNCKENLA
YFIGGFTGLGRAGDWSATRYEYVEALSVPEIDVSGIGRGNLLKFWFAQATAANYDVP
KFCAGFTNYKGFSAIYMTDLSLFRAYGAYSKPANDKLGSDFTFKRFDLGLIISAGF
KSNFNKIFKNRLALIISEGFTTFYNTKARFLIPEFLVNPIGPIDDTEITPTPPAD
LSAQGLEASAAKNSABQARAGAEAKPEKSKTDSVERWSILRSANALMSLQADKLGI
SNSNSESRSRSDVSTDTATPPPTPPEDDYTKQAATYDTFTSTSLADIQAALVS
LDQATYNKDTALQDEETAIAAEATFKNADAVKVGQAITYELAKYSDNOAIIILSLGKL
TSDFLQALALQSVNPKAAELLKEMQDNPVVPCKTPAIAYQSLVDOTATATQIEKL
GNAIPDAYQONASCVAENAKSNSNISLDSAKAALATAKTQIAEAQKFFPDSPILO
EAEQWIOIEKDLNKTIPADGSDVDPNGFTTVGSGKQDSSGISIRVSMILLDDAENETA
SILMSGSNPMHFNFTNQAQAQOELAAQAPAAKAAAGD3AAAAALDAQVLEAAAL
GRAGQGGTILNALGOTASAAVYSVPIPLQQVL"
659 a 435 c 480 q 574 t
BASE COUNT

```

Query Match	98.6%	Score 1369.4	DB 1	Length 2148	
Best Local Similarity	99.6%	Pred. No. 0			
Matches 1373	Conservative 0	Mismatches 6	Indels 0	Gaps 0	
QY	2	tgtttaatcctatgttccaggtccctatacagaacacagaaacagacacacctccgcagatc	61		
Db	770	TGTTTAATCCTATTGGTCCAGGTCCCTATAGACGAACACAGAACGCACACCTCCCGCAGATC	829		
QY	62	ttctgtcctaaggtatggagcgcagtgacagcaataagagtcgcggaagctcaagaataag	121		
Db	830	TTTTCTGCTCAAGGATTGGAGCGCAGTGCACAAATAAGAGTGCGRAGCTCAAAGAATAG	889		
QY	122	caagtcgcggaagctaagcctaaagaatctaagaccgattctgtagagcagatgagcaatct	181		
Db	890	CAGTGCAGGAAGCTAAGCCTAAAGATCTAAGACCGATTCTGTAGAGCGATGGAGCATCT	949		
QY	182	tgcgtttctgcagtgaaatgcgtcctcatagctcggcagataagctcgggtattgcctcttagta	241		
Db	950	TGCGTTCTGCAGTGAATGCTCTCATGAGTCTGCGAGATAAGCTGGGTATTTGCTTCTAGTA	1009		
QY	242	acagctgctgtcttactagcagatctgcagacgtggaactcaacgacagcgaccgcacct	301		
Db	1010	ACAGCTCGTCTTCTACTAGCAGATCTGCACAGCTGGACTCAAGCAGACGACCGCGACCTA	1069		
QY	302	cgcctctccaccacgcgtctgattataagactcaagcgcgaacacagcttcagatacta	361		
Db	1070	CGCTCTCTCACCCACCGTTTGATGATTATAGACTCAAGCGCAACAGCTTAGCATACTA	1129		
QY	362	tccttaacctcaacatcactagctgcacatacagcgtcgtttgggtgagctccacagatgctg	421		
Db	1130	TCTTTACCTCAACATCACTAGCTGACATACAGGCTGCTTTGGTGAGCCTCCAGGATGCTG	1189		
QY	422	tcactaatataaaggatacagcggctactgatatgaggaacccgcaatcgctgcggagtggg	481		
Db	1190	TCACTAATATPAAAGGATACAGCGGCTACTGATGAGGAACCGCAATCGCTGCGAGTGGG	1249		
QY	482	aaactaagaatgcgcgtgcgaattcaagttggcgcgcgaattacagaaattagcgaaatgtg	541		
Db	1250	AAACTAAGATGCGGATGTCAGTTAAAGTTGCGCGCAAAATTACAGAAATTAGCGAAATATG	1309		
QY	542	cttcgcgataaccacgagattcttgactctttaggtaaactgaactctcttcgcacctctac	601		
Db	1310	CTTTCGGATAACCAAGCGATTCTTTGACTCTTTTAGGTAACCTGACTTCTCTTCGACCTCTTAC	1369		
QY	602	agactgctcttcccaatctgtacgaacaaataacaaagcagctgagcttcttaagaga	661		
Db	1370	AGGCTGCTCTTCTCCAAATCTGTAGCAAAACAATAACAAAGCAGCTGAGCTTCTTTAAAGAGA	1429		
QY	662	tgcgaataaccacccagtagtccccagggaacacgcgctgcaattgctccaatcttttagttgatc	721		
Db	1430	TGCAAGATAACCCAGTAGTCCCAGGGAACGCCTGCAATTGCTCAATCTTTAGTTGATC	1489		
QY	722	agacagatgctacagcgcacacagatagagaaagatgcgaattggggatgcataatt	781		
Db	1490	AGACAGATGCTACGCGCACACAGATAGAAAGATGGAATGGCATTTAGGGATTCATATT	1549		



NSISNIDSAKALATATAKTOIAEAKKPPDPILOEABOMVIOAEKDLNKNIKPADSDV  
PNPPTTVGSKQSGSSISIRVSNLLDDAENEWASILMSGFROMINMENTENPDQAA  
QOEIAARAKARAGDDSSRAAALADAKALEALGRAGQOQILNALGOIASAAVSA  
GVPPAAASISGSSQYLYKSTGSDYKTOISAGYDAYKSINDAYGRANDATROVI  
NNVSTPALTRSPRPARTEARPEKTDQALARVISGNSRTLDGYISOVSALQSYMQLIQ  
SNPOANNEETROKLTSAVTRPPPOFGYPYVQLSNDSTQKFLAKLESFAEGSRTAAEIK  
ALSFETNSLFIQOVLVNIISLYSGLQ"

BASE COUNT 672 a 461 c 471 g 552 t  
ORIGIN

Query Match 97.9%; Score 1359.4; DB 6; Length 2156;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
Qy 1 atgggtaatcctattgttcaggctctatagacgaacacagacacacccctcccgagat 60  
Db 101 ATGGTTAATCCTATTGGTCCAGGCTCTATAGACGAACAGACGACACCTCCCGCAGAT 160  
Qy 61 ctttctgtcctaagattggaggcgaagtgcagcaataagagtcggaagctcaagaata 120  
Db 161 CTTTCTGTCTCAAGATTGGAGCGGAGTGCAGCAATAAGAGTGGCGAAGCTCAAGAATA 220  
Qy 121 gcagggtgcggaagctaaagcctaaagaattcgaagaccgattctgttagagcgatggagcatc 180  
Db 221 GCAGGTGCGGAAGCTAAGCCTAAAGAACTTAAGACCGATTCTGTAGCGGATGGAGCATC 280  
Qy 181 ttgcgtctgcagtgtaactctcctatgagctctgagctgagcagtaagctggttattcttctagt 240  
Db 281 TTGCGTCTGCAGTGAATCTCTCATGATGCTGGCAGATAAGCTGGGTATTGCTTCTAGT 340  
Qy 241 aacagctctcttctactagcagatctgcagcgtggactcaacacagcagcagcagcact 300  
Db 341 AACAGCTGCTCTTCTACTAGCAGATCTGCAGACGTGGACTCAACACAGCAGCCGACCT 400  
Qy 301 acgcctctcccccacgctctgatattataagactcaagcgcgaacacagcttcacgatact 360  
Db 401 ACSCCTCTCCACCCACGTTTGATGATTATAAGACTCAAGCGCAACAGCTTACGATACT 460  
Qy 361 attttacatcaactacatagctgacatacagggctgttggtagcctccagatgct 420  
Db 461 ATCTTTTACCTCAACATCACTAGCTAGCATACAGGCTGCTTTGGTAGCCTCCAGGATGCT 520  
Qy 421 gtcaactaataagatacagcggctactgatgaggaacccgcaactcgcctcggagtg 480  
Db 521 GTCRCRTAATAAAGGATACGCGGCTACTGTATGAGGAACCGCAATCGCTCGGAGTGG 580  
Qy 481 gaaactaagaatgccgatgcaattaaagttggcgogcaaatcacagaattagcgaatat 540  
Db 581 GAAACTAAGAATGCCGATGAGTAAAGTTGGCGCGCAAAATACAGAAATAGCGAAATAT 640  
Qy 541 gcttcggataacaaagcattcttgaactctttaggttaactgaactccttcacacttta 600  
Db 641 GCTTCGGATAACCAAGCGATCTTTGACTCTTTAGGTAAACTGACTTCTTCGACCTCTTA 700  
Qy 601 cagactgctcttctccaatctgtagcaacaataaacaagcagctgagctcttcaagag 660  
Db 701 CAGGCTGCTCTTCTCCAATCTGTAGCAACAATAACAAGCAGCTGAGCTTCTTAAAGAG 760  
Qy 661 atgcaagataaccagtagtcccgagggaacccctgcgaattgctcaacttttagtgat 720  
Db 761 ATGCAAGATAACCGCAGTAGTCCCGAGGAAACGCCCTGCAATTTGCTCAATCTTTAGTTGAT 820  
Qy 721 cagacagatgctacagcacacagatagagaagatggaatgcgattggagtgcatat 780  
Db 821 CAGACAGATGCTACAGCGACACAGATAGAGAAAGATGGAATGCGATTAGGGATGCATAT 880  
Qy 781 ttgcagcagacagcagtagtgagctgtagaataatgctaaatcttaataacagtagtaagc 840  
Db 881 TTTGCAGGACAGAACGCTAGTGGAGCTGTAGAAAAATGCTAAATCTTAATAACAGTAAAGC 940  
Qy 841 aacatagattcagctaaagcagcaatcgctactgtcttaagacacaaatagctgaagctcag 900

Db 941 AACATAGATTACGTAAGCAGCAATCGCTACTCTTAAGACACAAATAGCTGAAGCTCAG 1000  
Qy 901 aaaaagttcccccactctccaattcttcaagaagcggaaacaaatggtaatacagcgtcag 960  
Db 1001 AAAAAGTTCCTCCGACTCTCCAATTCTTCAAGAAGCGGAACAAATGTAATACAGGCTGAG 1060  
Qy 961 aaagatcttataaataatcaaacctcgagatgggttcttgatgttccaaatccagaactaca 1020  
Db 1061 AAAGATCTTAAATAATCAAACTGCAGATGGTCTGATGTTCCAAATCCAGGAACCTACA 1120  
Qy 1021 gttggaggtcccaagcaacaagaagtagtattggttagtatttcgtgtttccatgcttata 1080  
Db 1121 GTTCGAGGCTCCAAGCAACAAGGAAGTAGTATTGGTAGTATTCTGTTTCCATGCTGTTA 1180  
Qy 1081 gataatcctaaatgagacgccttccattttagtctggtgttctgagatgattcac 1140  
Db 1181 GATGATGCTAAAATGAGACCGCTTCCATTGTTGATGCTGGTTCGTCAGATGATTAC 1240  
Qy 1141 atgttcaatcgggaaatcctgattctcaagctgcccacacagagagctcgcagcacaagct 1200  
Db 1241 ATGTTCAATACGGAATAATCCTGATTCTCAAGCTGCCCAACAGGAGCTCGCAGCAACAGCT 1300  
Qy 1201 agagcagcgaagccgctggagatgacagtgctgctcagcgcgtcgcagatgctcauaaa 1260  
Db 1301 AGAGCAGCGAAGCCGCTGGAGATGACAGTGTCTGCGAGCGCTGGCAGATGCTCAGAAA 1360  
Qy 1261 gctttagaagcggctctaggttaaagctgggcaacaacagggcactactcaatgctttggga 1320  
Db 1361 GCTTTAGAGCGGCTCTAGGTTAAGCTGGGCAACAACAGGGCATACTCAATGCTTTAGGA 1420  
Qy 1321 cagatcgcttctgctgtgtgtgagcgcaggag-tcctcccgctgcagcgaagtctcatg 1379  
Db 1421 CAGATCGTCTCTGCTGCTGTTGTGAGCGCAGGAGTGTCTCCGCTGCGAGCAAGTTCTATA 1480  
Qy 1380 g 1380  
Db 1481 G 1481

RESULT 4  
AE001654/c  
LOCUS AE001654 10421 bp DNA BCT 01-DEC-2000  
DEFINITION Chlamydia pneumoniae section 70 of 103 of the complete genome.  
ACCESSION AE001654 AE001363  
VERSION AE001654.1 GI:4377031  
KEYWORDS  
SOURCE Chlamydophila pneumoniae CWL029.  
ORGANISM Chlamydophila pneumoniae CWL029.  
REFERENCE 1 (bases 1 to 10421)  
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE 2 (bases 1 to 10421)  
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.  
TITLE Direct Submission  
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
FEATURES  
source  
/organism="Chlamydophila pneumoniae CWL029"  
/strain="CWL029"  
/db\_xref="taxon:115713"  
133..2751  
/gene="Cpn0727"  
133..2751  
/gene="Cpn0727"  
/codon\_start=1  
/transl\_table=11  
gene  
CDS



```

/product="CT619 hypothetical protein"
/protein_id="AADI8866.1"
/db_xref="GI:4377032"
/translaton="MSFTSQNRRLRTISGESTRIIKDHYKSGFDPRSPVAINLEELN
SGIYALRLMLAQSENTAAALNPNTTPTSTWTDYKHSRPOASPRAPSQTPT
DIVSAALALVLVDGGLAEVASVTEIDIGALSTSTYRQLMASYLGLTTLTAQEK
VFGSWYNGWIKMAAAYQELAEYSLGTAVTSDQIKAYLQTRGNEFKATRHFFHNGD
QMTQFNETVFGNLTITANGAIOPLDGLGFIREAMTVGTVEADYVNAQRIILNEFNTA
LPHQEAELQIAELQKADDPGKASFENRKFVAAMTISESLGDALISMLNSQ
ATKQEAELKPLIEINENNAALNSLLOITNEFTSTVYVYSLSSVLSVQSGQNL
AGDYETLLAAAREERYIYRTARKQAKINLYNGLLQINSLPGATSAQKQEMNL
LYQISLSTVNLQTLVLESLLAGLKLQTTSSNNKYDKSEFKIESFDDWIPTLALESF
LTSGFFNSATGGLGFLFTQVSDQOQYTSOGQTOQLNQNMQTTIOEWTLVSTMQ
VLNGLISLAGAIYSN"
complement(3078..5033)
/gene="CpN0728"
CDS
complement(3078..5033)
/gene="CpN0728"
/codon_start=1
/transl_table=11
/product="CHLPN 76 kDa Homolog_1 (CT622)"
/protein_id="AADI8867.1"
/db_xref="GI:4377033"
/translaton="MWPVIGPGPIDETERTPPADLSAQGLEASAANKSABAQRIAGAE
AKPKESKTSVERMSILSRNALMSLADKLGIASSNSSSTSRSDVDTTATATPP
PPTFDYKTAQAYDTIFTSTLADIAQLVSLQDAVTNFKDQATDEETAAAEW
ETKNADAVYQAQGTTELAKYASNOAILDSGLKLTSPDLQALQLQSVANNAAEEL
KEMQNDVPVPGKPAIAQSLVDOTDATOIEKDGNAIRDAYFAGONASAVENAKSN
NSLSDSAKATATAKTOIAEAQKFPDPSILQEAQOMVIAEKDLNKKPADGSDV
PNQGTTVGSKQSGSGSIRVSMILLDAENETALMSGRFQMHMFNTENPDSQAA
QOEIAQAARAGADSDSAAALADAKALEAALGKAGQOQGLNALGOIASAAVWSA
GVPPAAASSIGSSVKLYKTSKTSQDYKTOISAGYDAYKSINDAYGRARNDRDVI
NYSVTPALTRVPRARTEARGPKTDQALARVISGNSRTILGDVYSQVSAQSQVQITQ
SNPOANNEERIKLTSVTKPPOFGYPYVOLNSDSTQKFTAKLESFPAEGSRTAIEK
ALSPETNSLFIOQLVNLGSLYSGYLQ"
complement(5145..6413)
/gene="CpN0729"
CDS
complement(5145..6413)
/gene="CpN0729"
/codon_start=1
/transl_table=11
/product="CHLPN 76 kDa Homolog_2 (CT623)"
/protein_id="AADI8868.1"
/db_xref="GI:4377034"
/translaton="MKKQVYQWLASVLLALTISGYAELPLSEQKVKSHYTTLLDEVK
DYLKRGFVETRQDQVLRAGDVRARMLYFREDIKNPSDKDKYNPLPVRNRYSEFYL
YIDYRAERNLWLSKMNMTAAGGNTAAGVDINRAFLGYRFYNPRTDFFMEIGRS
GLGLDFSEYQFOSNFDGLHIYMTRELSKDYPYQVIVHGGPFVYVNMKTKHYANVVEGI
LNRLPKQFFVKCSVVDNTEVPSETSTTEKAAANAMKYKCVQWQLVKKSQVPMWING
QKRPYLDIGAFNLAKATKTLNGKENLAWFIFGTLGLGRKAGDWSATVRYEYEA
LSVPEIDVSGIGRNLLKFWFAQIAAANYDPKEANGFTNYKGSALYMYGDTLSLSP
AYGASXPANDKLGSDFTFRKFDLGIISAF"
complement(6516..8159)
/gene="mvN"
CDS
complement(6516..8159)
/gene="CpN0730"
complement(6516..8159)
/gene="mvN"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AADI8869.1"
/db_xref="GI:4377035"
/translaton="MSRKDNVSLARSIFNLSGTFCSRTIGFIREIATAMATYFGADPI
VFAAFGLFVTFVFLKILGLLILEQAFIPHFELRAQSLDRAAFFFFRRSRLKNGSI
IFTLLEAVLVLYVVEGTYDMILLTLLPCGIFLMNMYNVALHLLHCENFFGVG
LAPVWNIWIFEVIAARHSDPREIRIIGLSVALVIGFFFWLITVPGVWFLEAKSP
POHDSVRALLAPLSGLITSSIFQLNLLSDICLARIYHEIGPILYLMYSKTIQOLPH
LFGFVGFVTLPLPAISRCVQRDHERGLKMKFVLTLTMSVMIIMTAGLLLLALPGVVR

```

BASE COUNT 3278 a 2158 c 1983 g 3002 t  
ORIGIN

Query Match	Best Local Similarity	Score	DB 1	Length	10421
Matches	1374	Conservative	0	Mismatches	6
Indels	1	Gaps	1		
Qy	1	atggttaatcctattggtccagggtcctatagacgaacacagacacacccctcccgagat	60		
Db	5033	ATGCTTAATCTATTGTGTCAGGCTCTATAGCAACAAACACGACACCTCCCGCAGAT	4974		
Qy	61	ctttctgctcaagattggagcgaagtcacagaataagagtcggaagctcaagaata	120		
Db	4973	CTTTCTGCTCAAGATTGGAGCGGAGTCGACCAATAAGTAGTGGGAAGCTCAAGAATA	4914		
Qy	121	gcaggtgcggaagctaaagcctaaagaattcttaagaccgattctgtagacgattgagc	180		
Db	4913	GCAGGTGCGGAAGCTAAGCCTAAGAATCTAAGACCGATTCTGTAGACGATGAGCATC	4854		
Qy	181	ttgcgtctgcagtgaaatgctctcatgagctgagcagtaagctgggtgattgcttctagt	240		
Db	4853	TTGCGCTTCTGCAGTGAATGCTCTCATGAGCTGCGAGATAAGCTGGGTATTGCTTCTAGT	4794		
Qy	241	aacagctctcttactagcagatctcagacgtggactcaagacagcagccacact	300		
Db	4793	AACAGCTCGTCTTCTACTAGCAGATCTCGACACCTGGGACTCAACGACAGCAGCCACCT	4734		







```

gene
HQGCCGEGCSQSPVRKFDSDSRGNKPFRAHLCEISENSESASCCLI"
2238. .2867
/gene="CP0013"
CDS
2238. .2867
/gene="CP0013"
/note="similar to SP:P81288; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S4"
/protein_id="AAF37909.1"
/db_xref="GI:7188952"
/transl_table="MARIYCGPKNVRARFGANIGFRSRNPLKKPKPGHGMQRKKK
SDYGLQEEQKQKACACYMIMKOLVKAFKEVHKQCNVQAMFLERCECLDNNVYRM
GFAKTIFAAQOLVAHGHLVNGRVRDRSFLRPGMOISLKEKSKRLQSKDALESKD
ESSLPYSISLDKKGCKGELLVSPQDQIEAQLPLPINISVVCFLSRT"
complement(2986. .3867)
/gene="CP0014"
complement(2986. .3867)
/gene="CP0014"
PID:405898 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="endonuclease IV"
/protein_id="AAF37910.1"
/db_xref="GI:7188953"
/transl_table="MKVLPPSPITLLGAHTSTAGGLKNAIYEGRDIGASTVQIFTAHQ
RQWRRAKKEVEDFKAALKEDTSLIMSHAGLYINPGADPVPVILEKSRIGIYQEL
DCLTIGSFVNFHFGAALKSKEDKMKIVSSFSQSPADFDSSPLVLLTLAGQGT
LNGFEEGLYVQNLKNQIPIGVCDTCHIFAAGYDITSPQWEDVINEFDEYVGLS
YRAFHLNDSMFLPGANKRHPALGEGYIGKESFKLMTDERTKIPKYLETPGPGEN
WQKEIGELLAKFSNRDS"
complement(4202. .4378)
/gene="CP0015"
complement(4202. .4378)
/gene="CP0015"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF37911.1"
/db_xref="GI:7188954"
/transl_table="MSFLIKPFPPLKDYLDISPYTSSEITIGGSYFLKNKASL
QSSTLRLSISILS"
4362. .6005
/gene="CP0016"
4362. .6005
/gene="CP0016"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37912.1"
/db_xref="GI:7188955"
/transl_table="MSRKDNVLSIRFIENLSTFCSTRITGIFREIATAMATYEGADPI
VAAPFLGFRVFLKRLTGLGLLEQAFIPHEFLRAQSLDRRAFFFRFRSLKFGSTI
IFLLIEAVLWVQLVVEGTYDMILLTLLPCGIFLMMYNVNGALLHCENKFFGVG
LAPVNNIMIFFVIAARSDPRRIIGLSVALVIGFFFWELITVPQWVKFLLEAKSP
POEHSVRALLALSLGLTSSIFQLNLLSDICLARYVHEIGPLIYMSKLYQLPIH
LFGFVTVLLPAISRVCQREDREGLKMKFVLTMTSMIIMTAGLLALLALPGVRV
LYEHLGFPQSAVAIVRLVRYGASIIIPMALVPLVSVLVAQRQYAVPLFGIGTALA
NIVLSVLGRWIKLVKDSGISATSIATWOLYFLWYSSKRLPMYSKLLWESIRRSK
VMGTMLACMITGLNLTOTTTIFLNLPLPLWPLUSSITAQIAFLSBSICFLAFL
FGAKLLRVEDLINLASFEYWRGQGLLQROHWQDQTN"
6108. .7376
/gene="CP0017"
6108. .7376
/gene="CP0017"
/note="conserved hypothetical protein; identified by

```

```

Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37913.1"
/db_xref="GI:7188956"
/transl_table="MKQVYQWLASVLLALTISGYAELPLSEOKVKSHYTTLLDEVK
DYLKRCGFVETRKQDGLRIAGDVRARMLVFEREDIKNPSDKYKPLPVARYSEFVL
YIDYRAERNLSSKMWTAAGTENTAGVVDINRAFLCYRFXKMPETRTDFMEIGRS
GLDLFESEVQFSNFDGLHIYTRRELSDYPIQVIVHGGPFVYVMMTKHYAWVVEGI
LNLKPKQFFVKCVSWDMNTFVPSSTTERKAATNAMKYCYVMQWLKGSQVFWING
QKPLLYLGAFALMNLAKATKTTLNGKENLAWFIGGTLGLLRKAGDMSATVRYEYEA
LSYPEIDVSGIGRNLKAFWFAQAIANYDPKEANGFTNYKGFSAIYMYGTTDSLSFR
AYGASKRPANDKLGSDFTFRKDLGIISAF"
7488. .9443
/gene="CP0018"
7488. .9443
/gene="CP0018"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37914.1"
/db_xref="GI:7188957"
/transl_table="MVNPIGPGPIDETPTPADLSAAGLEASAANKSAAQRTAGAE
AKPKESKTSVERWSILRSVNALMSLADKLGIASSNSSSTSRSDVDTATATPP
PPFTDDYKQAOATAYDTITFTSLADIQAALSLQDAVYNIKDTATDETAIAEAE
ETKNADVPVQKITEUKAYASQNOAILDSGLKTSFDLQALQLQSVANNKAAELL
KEMQDNVPVKGPTAIQASLVQDQDATQTEKDNARIDAYFAGQASGAVENAKSN
NSISNIDSAKAATATAKTAQAEAKKFPDPILOEAQPMVIAEKDLNKTIPADGSOV
PNPQTTVGGSKQOCSSTISRVSLMDDAENETASILMSGFROMIHMFENTENPDQAA
QOELAAQARAAGDSDAAALADAQALEAKGKAGQOQILNALGOLASAAVVA
GVPPAAASSIGSSVKLYKTSKTSQSYKYQISAGYDAYKSINDAYGRANDATDVTI
NNVSTPALTRSVPRARTKAPGPPQPGYIPYVLSNDSTQKTALESFAEGSRATAEIK
SNPOANNEIRKQLTSATVPPQPGYIPYVLSNDSTQKTALESFAEGSRATAEIK
ALSFTNSLFIQQVLVNIQSLYGLQ"
complement(9770. .12388)
/gene="CP0019"
complement(9770. .12388)
/gene="CP0019"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37915.1"
/db_xref="GI:7188958"
/transl_table="MSTFSIQNRLRTISGESTRIKLDHKYSGFDRSPVPAINLEELN
SGIYALRLNALQSENTNVAALINPNNTIPPTTSMTDYKHSRPOASPRAPSOTPT
DIYAAALALVLDIGLAELVASVTEIDIGALSTISTVRLMASYLGTLTLTAEQEK
VPSSSYVPSEKLLHVKQEKAEIOAKEEIKAVLEAKGVSTEEIEALKEYPDII
ADPFKEIEEPLHYTRAKVAPIQEMENAIQLLTPPALTPDNVNEVNGMTLSTLI
LQADDAIKQAPALGGQDEIITLQPLVPLDTFTTFAEFDLIYATQPLTASLKL
YLTDQIAEYRGKITKYQNSIQNLSEKRVNNRSMLETSQSMFOAQCNCFYTWIS
QANALNIAITNKYISAVLTSMEYGLGLCLSYMYERLADDERAIFDKSNCEYLPHI
VWGSVWNGWIAKMAAYELAEYSLGTAVTSDQIKAYLOTGRNEKATRHEFINIGD
OMQCPANVETGNCITPANGAIOPLDGGFIREAMTNNVTVEADYVSNARILNEPNTA
ATAHVLQLOIAELQKADDDPGKASFTENKFAVAAMITSESDALISMLNSQ
LPQKAEFLKPLIEEINFNNLAANLNLQITNEFTSYVYLSVSVQSVKSTQGNLF"
Query Match Score 1359.4; DB 1; Length 12548;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Qy 1 atgtgttaactctatgtccaggctctatagacgaacacagacacaccccccagat 60
|||||
Db 7488 ATGTTAATCTTATGTGTCAGGCTCTATAGACGAACAGACACACCTCCCGCAGAT 7547
|||||
Qy 61 cttctcgtcaagatgagcgagtcgacgaataagagtcggaagctcaagaata 120
|||||
Db 7548 CTTCTGCTCAAGATTGGAGGCGAGTGCAGCAATAAGAGTGGGAAGCTCAAGAATA 7607
|||||

```



Qy	121	gcaggctgcgaagc	taagcctaaagaatc	taagaccgattct	gtagagcgtagagc	atc	180
Db	7608	GCAGGTCGGAAGCT	TAAGCCTAAAGAACT	TAAGACCGGATCT	TGTAGACGATGGAGCATC		7667
Qy	181	ttcgcttctcagtg	aatgcctctcat	gagcttgcgcagata	agctgggtattgcctctcag	t	240
Db	7668	TTGGTCTCGAGTGA	ATGCTCTCATGAGCT	TGGCAGATTAAGCT	TGGGTATTGCTTCTAGT		7727
Qy	241	aacagctcgtcttct	actactagcagatct	gcagacgtggagct	caaacgacagcagccgac	cct	300
Db	7728	AACAGCTCGCTTCT	TACTTAGCAGATCT	GCAGACGTGGACT	CAACGACAGCAGCCGACCT		7787
Qy	301	acccctctccaccac	gctctgaatata	gaagactcaagcgcaaac	agctctcagatact		360
Db	7788	ACGCTCTCCACCC	ACGTTTATGATTATA	GAGCTCAAGCGCAAC	AGCTTTAGGATACT		7847
Qy	361	atcttaccctcaac	atcacctagctg	acatacacaggctg	cttggtagcctccagatg	ct	420
Db	7848	ATCTTTACCTCAAC	ATCACTAGTGCAT	TACAGGCTGCTTT	GGTGAGCTCCAGGATGCT		7907
Qy	421	gtcaataataaaga	gatacagcgctact	gatgaggaaaac	cgcaatcgctgcggagt	gg	480
Db	7908	GTCACTAAATATA	GAAGGATACGCGGCT	ACTGATGAGAAAC	CGCAATCGCTCGCGAGTGG		7967
Qy	481	gaaataagaatgcc	gatgcaattaa	agctggcgcaaat	tacagaattagcgaatat		540
Db	7968	GAACTAAGAATCC	CGATCGATTAA	GTTGCGCGCAAA	TTACAGAATTAGCGAATAT		8027
Qy	541	gcttcggataacca	agcgattcttgact	ctttagtdaaact	gaactccttcgcagctctta		600
Db	8028	GC TTCGGATTAAC	CAACGCGATTCT	TGACTCTTTAGTG	TAAACTGACTTCCTCTCGACTCTCTTA		8087
Qy	601	cagactgctcttcc	aatctgtacaa	acaataacaagcag	ctgagctctcttaagag		660
Db	8088	CAGCTGCTCTTCT	CCANTCTGTAGCA	ACAATATACA	AAGCAGCTGAGCTCTTTTAAAGAG		8147
Qy	661	atcgagaataacc	cgatagctcc	cagggaacacgct	gccaattgctcaactctttagttg	at	720
Db	8148	ATGCAAGATTAAC	CCAGTAGTCC	CAGGGAACCGCT	CAATTTGCTCAATCTTTAGTTGAT		8207
Qy	721	cagacagatctac	agcgacacagata	gagaagaatg	gaatggattgggagtcg	at	780
Db	8208	CAGACAGTCTTAC	CGACACAGATAG	AAGAAGATGGAAT	TGCGATTTAGGGATTCGATAT		8267
Qy	781	tttcaggacagac	gcgtagtgagctg	tagaaatgctaa	atcttaaacagataa	gc	840
Db	8268	TTTCGAGACAGAC	CGTAGTGAG	AAAAATGCTAAA	ATCTTAATACAGTATAAGC		8327
Qy	841	aacatagattcag	taaaagcagcaat	cgctactgcta	gagacacaaa	tagctgaagctcag	900
Db	8328	AACATAGATTTC	AGCTTAAGCAGCA	ATTCGCTACTGCT	TAGACACAAATAGCTTGAAGCTCAG		8387
Qy	901	aaaaagtctccc	gactctccaatt	cttcgaagcgga	caaaaatgtaata	cagcgctgag	960
Db	8388	AAAAAGTTCCCG	CACTCTCCAA	TTCTTCAAGAAG	CGGAACAAATGGTAA	TACAGGCTCAG	8447
Qy	961	aaagatcttaaaa	ataataaacct	gcagatgctctg	atgttccaaatcc	caggaaactaca	1020
Db	8448	AAAGATCTTAAAA	ATATAAACCTC	GCAGATGGTCTG	ATGTTCCAAATCCAGGAAC	TACA	8507
Qy	1021	gttgaggtccaa	gacaacaggaat	gattgttagtatt	cgctgtttccatgctg	tta	1080
Db	8508	GTTGGAGGCTTCA	AGCACACAGGA	AGTAGTATTGGT	TAGTATTCGTGTTTCCAT	GCTGTTA	8567
Qy	1081	gatgatgctgaaa	aatcctgatctc	aaagctgcccaac	aggagctgcgacagca	agct	1140
Db	8568	GATGATGCTGAAA	ATAGACCGCTT	CCATTTTGATG	CTCTGGGTTTCGT	TCAGATGATTCAC	8627
Qy	1141	atgttcaatac	gaaaatcctgat	ctcagctgcc	aaacaggagctgcgacagca	agct	1200
Db	8628	ATGTTTCAATAC	GGAANAATCCT	GATTCTCAAG	CTGCCAACAGGAGCT	CGACAGCAAGCT	8687
Qy	1201	agagcagcaaac	gcctgagatdac	atgctgctgc	gcagcgctgcgagat	tgctcacaaa	1260

Db	8688	AGACGACGAAGCCGCTGGAGTACAGAGTCTCTCGACGCTGGCAGATGCTCAGAA	8747
Qy	1261	gctttagaagcggctctaggttaaagctgggcaacacagggcatactcaatgctttggga	1320
Db	8748	GCTTTAGAACGGCTCTAGGTAAACCTGGGCAACACAGGGCATACTCATGCTTTAGGA	8807
Qy	1321	caagtcgcttctgctgctgttgtagcgaggag-tctctcccgctgcagcaagtctatg	1379
Db	8808	CAGATCGTCTTCTGCTGTGTGTGAGCGCAGGAGTCTCTCCGCTGCAGCAAGTCTATA	8867
Qy	1380	g	1380
Db	8868	G	8868
RESULT	6		
AP002547/c			
LOCUS			
DEFINITION	AP002547 300550 bp DNA BCT 08-DEC-2000		
	Chlamydothila pneumoniae J138 genomic DNA, complete sequence,		
	section 3/4.		
ACCESSION	AP002547 AB033786 AB033787 AB033816 AB033817 BA000008		
VERSION	AP002547.2 GI:10176693		
KEYWORDS			
SOURCE	Chlamydothila pneumoniae J138 (strain:J138) DNA.		
ORGANISM	Chlamydothila pneumoniae J138		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothila.		
AUTHORS	1 (sites)		
	Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K.,		
	Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.		
TITLE	Comparison of whole genome sequences of Chlamydia pneumoniae J138		
	from Japan and CWL029 from USA		
JOURNAL	Nucleic Acids Res. 28 (12), 2311-2314 (2000)		
MEDLINE	20330349		
REFERENCE	2 (bases 1 to 300550)		
AUTHORS	Shirai, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-JUL-2000), to the DDBJ/EMBL/GenBank databases.		
	Department Shirai, Yamaguchi University School of Medicine,		
	Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi		
	755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,		
	tel:81-836-22-2227, fax:81-836-22-2415)		
	On Sep 15, 2000 this sequence version replaced gi:6172398		
	gi:6172300 gi:6172396 gi:6172398 gi:8978889.		
COMMENT	AB033786-AB033787, AB033816-AB033817: Submitted (25-Oct-1999).		
FEATURES	Location/Qualifiers		
source	1. 300550		
	/organism="Chlamydothila pneumoniae J138"		
	/strain="J138"		
	/db_xref="taxon:138677"		
gene	/notes="synonym:Chlamydia pneumoniae (strain J138)"		
	complement(21..794)		
	/gene="dapF"		
CDS	complement(21..794)		
	/gene="dapF"		
	/codon_start=1		
	/transl_table=11		
	/product="diaminopimelate epimerase"		
	/protein_id="BAA98725.1"		
	/db_xref="GI:8978890"		
	/translation="MAFYSPSTISKVFIYSGAGNRFLIGTLPEVDDVRFLLCOFTRVD		
	GFLLXPSSCAQAQLIFNSDGRPTMCGNGLCAHAHLASOKGSDISYSTOSGLYS		
	GYFYSDRWLVDMTLADSRPVRHRLSRPDLPKEVVCIHGTGPHAVVILPILSLDVL		
	SILGLPRLHYQTSYDPGVNVNFVILGHQCLRVRTYERGVEGTAACGTGALASALIV		
	SNSYGNKESIQIHTWGGELMTVSQNRGVLYQSVTRDL"		
	complement(763..1338)		
gene	/gene="clpP_1"		
	complement(763..1338)		
CDS	/gene="clpP_1"		
	/codon_start=1		
	/transl_table=11		
	/product="CLP protease"		
	/protein_id="BAA98726.1"		



```
/db_xref="GI:8978891"
/translation="MAGEVHKLRDITTEKELLEARRVFFSEPVTEKSASDAIKKLWYL
ELKDPGPPIVINSWPGSVNDAGFAVNDQIKMLTSPVTVVTGTLAASGWSVLSCAAP
GRRATPHSRIMIHQPSIGGPIQTQANDLDIHAREILKTKARIIDVVEATNQPRDII
EKAIDRMWMTANAKAFGLDGLDGILFSPNDL"
gene
CDS
complement(1358. .2917)
/gene="glya"
complement(1358. .2917)
/gene="glya"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/translation="MLKVEKFKFAIVEIFTKVAVVVLHKLFLNAGSKGKGOSLAS
TAYLAADHLNAPSGTIGERIIDELKSORSHLMKASENYSLSVOLAMGNLLTDKVC
EGSPKFRFYSCENVDALNEWCETAKELFADACVQPHSGADANLAVNAILTHKV
QGPVSKLGYKTYVNELEETTLIKAESNCSCIGPSLNSGGHLTHGNVRINWMSKM
RCFPYDNPDTCEFDYAEISRLAKPKVLIIAGYSSYSRRLNFAVLKQIAEDCGSVL
WDMHAFGLPVAGVGFVDEENPIYADIVTTTHTKLRGPRGLVATREYESTLNKA
CPLMMGGPLHIVIAAKTVALKALSVDFKKYAHQVNNARLARLRFSLHGLRLLTGGT
DNHMMVIDLSLIGISGKTAEDILSSVGLAVNNSLPSDAIGKWDTSGLRLGTPLTIL
GKGDMEEEVADIIVKVLNRLNRLSCHVEGSKKNKGELPEAIAOEARDRVNLLLRFP
LYPEIDUEALV"
3107. .3778
/gene="CpJ0522"
3107. .3778
/gene="CpJ0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/translation="MTLYLGLNOKTARKYOAHLPLILTLPYAKSTONKRALQFLPQ
ATHVILSPSSTHFLSMTLSLLKATLKTKTYICIGESTKERLUSFLGQKVVYATQ
LSKYGIPPLQALPSSAKILYHSSLARPIREFLYNNRFTFSPHYHYVPRKLKNNI
LSKQKIIFTSPTVRAPAKIFPRFPPEKTYWCQGRMTLQEQFKFSQKQVSLLETLGK
SRTSP"
3843. .4175
/gene="CpJ0523"
3843. .4175
/gene="CpJ0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/translation="MASSATGFCGTAPSLFPATPRYNNKALFVTTIALVWIAL
IATTIAGLCIHLPCSFIFLTAIPLIFISRYICSHYARNVYIALDVVPDHSKLODMRS
HSPIFSFR"
4223. .5302
/gene="CpJ0524"
4223. .5302
/gene="CpJ0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/translation="MSGPSRTFESSQSVLSVPRDKETAPKKQTTIAKISTAILASL
ALGAVAGISLITVILGNPFLALLITLALFVTVFLVYHMTKSVSSNMWQLEONFK
PLGAWQKNKYDCYSNEMQFYNNHLNPKFVAIQTDASQFPQFTLGLRVIERNQST
GIFNPGVTNLIDNTAINLSTILYSLTKDKSVWDTCKQREGGAKGDDPFSPTVEYR
VKLNEADQTFNLNLSAEKKSILPTFLGHVCGPKSEELPNQOEYVROALLAYENCL
KAATESHAALVALPFTSVKVEVPEEILPKESTFYWDNQTOAFCKRALLDAIQNTALR
YQPSRLVILQDPNFTESQSRSE"
5642. .6406
/gene="CpJ0525"
5642. .6406
/gene="CpJ0525"
/codon_start=1
/transl_table=11
/product="CT398 hypothetical protein"
```

```
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/translation="MHDALLSILAIQELDIKIMRLMRVKKHKEQKELAKVOSLKSDIIR
KVQKLEMLNLTQIRDENRNIQIEISEQINKENQAQAAVKMKDEFNALTOEMTTANK
ERRSLEHQLSDMLDKQAGEDLIVSLKESLASSTSSVIEKEFESIKKINEGKAL
LEQRTLKHATNPPELLSIYERLLNNKRDVVVPTIENRVSCGCHIVLTPOHLENLVRRKD
RLIFCEHCSRILYQESQVNAQENSTAKRRRRRAV"
complement(6827. .7816)
/gene="kpsF"
complement(6827. .7816)
/gene="kpsF"
/codon_start=1
/transl_table=11
/product="Guto/KpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/translation="MPSPMISTDVQCQILGKQKQKAVDFFFOAQFQKQAMQAEKILGH
SGWYFFGKGVKGCVARKLKVLATLSERALLQSPVOLLHGDGLVSPGDIVCLFSKS
GETOELLDTVPLKSRRAILVAITSMYPSNLAAISDLVVLPSVAELDPFLIPTNST
TCOMIFGDFLHMLLFHSRGVSLSTYKNHPSGOVMKANGVKDFMPKTPVFPCHLG
DKVSFLESFSAVCGGCVCIVDPQFLMGIFPDGDLRSLASYGGEVLSLSLEKVMTA
NPRCITEDSDIAIQLMESSPPAVLPVLNENRHVTLGLLHMTLAKAGLL"
complement(7843. .9024)
/gene="sucB_2"
complement(7843. .9024)
/gene="sucB_2"
/codon_start=1
/transl_table=11
/product="dihydrolipoamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/translation="MIFEPFPKIGETSSGGSIVRWLKNLGDHVARDEPLIEVSTDKI
ATELPSKAGLRFVFCYNEGDEVAGDVGLIELEEISEADDESTSCPLTSCETKSEA
GSSSSWFSFPAVLSLAQREGIGLDNLQK IAGTKGGRVTRQDLEAVISEQQSIPDE
IFQGEVNRPMSPURRAIASLSLSSDSDEVPHASLVVDVDDTLNMLNLSGERGFLDTH
GVKLTITTSFVQCLAAQLRQFLLNGSLDGTIIVMKKSVNVGVAVNLNKEGVVPIVH
NCODRGLVYSKLAADLSSRARLNKLDPSVQDGSVTVTFNGMTGALIGMPIIRYPEV
AIGIGTITQKRVVVRDDDSLAIKRMVYVTLTFDHRVLDGIYSGSEFLTSLKNRLESVTM
G"
complement(9038. .10282)
/gene="glut"
complement(9038. .10282)
/gene="glut"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/translation="MKLWMKIFGLFVGVTLGLVLEDKAIFPKPIGDIFLNLMSVVY
PLVFCSMVLGIASIDMKKLGRIKSVGLYLTALAIVIGLCFAWIFSPGNGCDEFA
QAQSMDSAVTVIDSNKTAAYFLSIIAOVFPSPNVPVRSFAEGNILOIITIFLFLALRL
SGGERPVERFIDGSEITMLRWNNKMSFAPYGVGASMAWISGNHGLVWLQGLKFII
AYYLACLFHATLVFGLVRFCCKMSFSKFLSSMDAISCASVSTASSATLPVTRCVS
KNLGVSAEVSGLVPLGATVNNGTAIPOGMAAVFIAQAYNCPULSSLLLVVVTATF
SAGSAGVPGGEMITLGSVSLASVGLPIQGIATLAIAGIDRLRIDVGTPTMNLGDVAVATY
VAGGELSPPYSIKQESVET"
complement(10282. .11379)
/gene="ycaH"
complement(10282. .11379)
/gene="ycaH"
/codon_start=1
/transl_table=11
/product="ATPase"
/protein_id="BAA98735.1"
/db_xref="GI:8978900"
/translation="MKRRFPTLFLFRRVVTIAISLEGILGWGLGSLLSKVFAFLVA
CWNRFSTVPYRSTVSVISGVNIVVGAGKTPTVTLWLAELRLRGYSCGLVSRGYSKQ
SSRQKLTVDVDSKVSASVYGDPEPLMAEKIPESVWVHKDRRISAAARAEKFGILL
DDGLQYRKLHKDVEIAVNGODPLGRDRDFPLRLKTVDAITVNGGKEAG
TVKVRSNAPDIFVKPTIASVVVTHNGERIPKEALRELVRGVFCGLGPGQFLANLRE
EGIHLLGKYLPLPDHAAITKKELNYFCQOMAMROGQGLLCTEKDSVKLPRLSGEVSLLP
TAKVEMRLSVNQDDTSLNLNMQIEQIHNRGN"
```



Query Match	97.9%	Score 1359.4	DB 1	Length 300550
Best Local Similarity	99.5%	Pred. No. 0		
Matches 1374	Conservative	0	Mismatches	6
Indels	1	Gaps		
QY	1	atggttaactctattggtccaggtctctatagacaaacagacacacacccgcagat	60	
DB	217598	ATGGTTAATCCCTATTGGTCCAGTCTCTATAGACGAAACAGAACACACCTCCGCGAGAT	217539	
QY	61	ctttctgctcaaggattgaggcgagtcgacaaataaagatgcgggaagctcaagaata	120	
DB	217538	CTTTCTGCTCAAGGATTGGAGCGAGTGCAGCAATAAGAGTCGGGAAGCTCAAGAAATA	217479	
QY	121	gcagggtcggaagctaagcctaagaactaagaccgattctctagagcgatggagcgc	180	
DB	217478	GCAGGTGCGGAAGCTAAGCCTTAAGANCTAAGACCGATTCTGTAGAGCGATTGGAGCATC	217419	
QY	181	ttagcttctgcagtgaaatgctctcatgagctggcagataaagctgggtattgctctagt	240	
DB	217418	TTGCGTTCTCGAGTGAATGCTCTCATGAGCTGGCAGATAAGCTGGGTATTGCTTTCAGT	217359	
QY	241	aacagctcgcttctactagacatctgcagactggagctcaacgacgagacgcacct	300	
DB	217358	AACAGCTGCTGTTCTACTAGAGATCTGCAGACGTGGAGCTCAACGACAGCGACCGACCT	217299	
QY	301	acgctctccaccccgctctgatgattataagactcaagcgcaaacagctctacgatact	360	
DB	217298	ACGCTCTCTCCACCGTTTGATGATTATAAGACTAAGCGCAACACAGCTTACGATACT	217239	
QY	361	atctttacctcaacataactagctgacatacagggctgcttggtagagctccaggaatgct	420	
DB	217238	ATCTTTACCTCAACATCACTAGCTGACATACAGGCTGCTTTGGTAGCGCTCCAGGATGCT	217179	
QY	421	gtcaactaataaagatagacagcgctactgatgaggaacgcgaatcgctgcggagtg	480	
DB	217178	GTCACTAATATTAAGGATACAGCGCTACTGATGAGGAACCGCAATCGCTGCGAGTGG	217119	
QY	481	gaaactaagaatccgatgcaattaaagtggcgcgcaaatcacagaattagcgaatat	540	
DB	217118	GAAACTTAAGAAATCCGATGCAGTTAAAGTTGGCGCGCAAAATTACAGAAATTAAGCAAAATAT	217059	
QY	541	gcttcgataaccaagcattcttgactctttagttaaactgactctctcgacctcta	600	
DB	217058	GCCTCGGATACCAAGCGATTCTTGACTCTTTAGGTAAACTGACTTCTCTGACCTCTTA	216999	
QY	601	cagactgctcttccaatctgtagcaaacataacaagcgagctgactctttaaagag	660	
DB	216998	CAGGCTGCTCTTCTCCAACTGTGTAGCAACAATAACAAAACGAGCTGAGCTTCTTAAAGAG	216939	
QY	661	atgcaagataaccagtagtccacagggaaaaacgctgcaattgctcaatctttagtgat	720	
DB	216938	ATGCAAGATACCCAGTAGTCCACGGGAAACCGCTGCAAAATTGCTCAATCTTTAGTTGAT	216879	
QY	721	cagacagatgctacagcgacacagatagagaagaatggaaatgcgattggggaatgcat	780	
DB	216878	CAGACAGATGCTACAGCGACACAGATAGAGAAGATGGAAATCCGATTAGGAGATCATAT	216819	
QY	781	tttcagagacagaacgctagtggagctgtagaaaaatgctaaatcttaataacagtaaacg	840	
DB	216818	TTTGCAGGACAGAACGCTAGTGGAGTGTAGAAAATGCTTAAATCTAATAACAGTATAAGC	216759	
QY	841	aacatagattcagctaaagcagcaatcgctactgtaagacacaataagctgaagctcag	900	
DB	216758	AACATAGATTTCAGCTAAGCAGCAATCGCTACGTGTCTTAAGACAAATAGCTGAAGCTCAG	216699	
QY	901	aaaaagtctcccgcactctccaatcttcaagagcggaacaaatggtaacagcgctgag	960	
DB	216698	AAAAAGTTCCCGACTCTCCAAATCTTCAAGAAGCGGAACAATGGTAATACAGGCTCGAG	216639	
QY	961	aaagatcttaaaaatatcaaacctcgagatgggtcttgatgttccaaaatccaggaactaca	1020	
DB	216638	AAAGATCTTAAAAATATCAAACTCGAGATGGTGTCTGTGATGTTCCAAATCCAGGAATCTACA	216579	
OY	1021	gttgagagctccaaagcaacgaagaatgattggtaatgctatgcttctccatgctgtta	1080	

Db	216578	GTGGAGGCTCCAAAGCAAGAAAGTAGTATTGGTAGTATTGGTGTTCCTCATGCTGTTA	216519
QY	1081	gatgatgtgaaaaatgagaccgcttccattttgatgtctgggttttcgtagatgtatcac	1140
Db	216518	GATGATGCTGAAAATGAGACCGCTCCATTTTGTCTGGGTTTCGTGAGATGATTCAC	216459
QY	1141	atgttcaatcggaaaatcctgattctcaagctgcccacaacagagagctcgacgacaagct	1200
Db	216458	ATGTTCAATACGGAATACTGATTCTCAAGCTGCCCAACAGAGAGCTCCGACGACAAGCT	216399
QY	1201	agagcagcaaaacccgctgagatgacagtctgctgctgcagcgtgcgagatgctcagaaa	1260
Db	216398	AGAGCAGCAAAACCGCTGAGATGACAGTGTCTGCTGCAGCGCTGGCAGATGTCAGAAA	216339
QY	1261	gctttagaacggccttagtaagctgggcaacaacaggggcatactcaatgctttggga	1320
Db	216338	GCTTTAGAAGCGGCTCTAGTAAAGCTGGCAACACAGGCGATCTCAATGCTTTAGGA	216279
QY	1321	cagatcgtctctgctgctgtgttgagcgcaggag-tctctcccgctgcagcaagtctctatg	1379
Db	216278	CAGATCGCTTCTGCTGCTGTGTGTGAGCGCAGGAGTTCTTCCCGCTGCAGCAAGTCTATA	216219
QY	1380	g 1380	
Db	216218	G 216218	
RESULT	7		
AX045135	AX045135	1456 bp	DNA
LOCUS	Sequence	5 from Patent	WO0066739.
DEFINITION	AX045135		
ACCESSION	AX045135		
VERSION	AX045135.1	GI:11343734	
KEYWORDS			
SOURCE	Chlamydomophila pneumoniae.		
ORGANISM	Chlamydomophila pneumoniae		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.		
AUTHORS	Murdin, A.D., Omen, R.P., Wang, J. and Dunn, P.		
TITLE	1 (bases 1 to 1456)		
JOURNAL	Thereof		
FEATURES	Patent: WO 0066739-A 5 09-NOV-2000;		
source	Aventis Pasteur Limited (CA)		
	Location/Qualifiers		
	1..1456		
	/organism="Chlamydomophila pneumoniae"		
	/db_xref="taxon:83558"		
CDS	101..>1452		
	/note="unnamed protein product"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAC17379.1"		
	/db_xref="GI:11343735"		
	/translation="WNPIGPGIDETERTPPADLSAQGLEASAAKNSAPAOPIAGAE		
	AKPKESKTSVRSWISILRSVNAIWSLADKLGIASSNSSSTSRKADVDSTTATATP		
	PPFTDDYKTAQATYDTITFTSLADIQAALVSLQDAVTNIKDTATDDETAIAEAP		
	ETKNADAVKGAQITAYKASDNQAILDSGLKTFDLQAALLSVAANNKAELL		
	KEMQNDVPVQKTPATAQSLVDQTDATATIEKDGNARIDAYFAGONSAGVANNAKSN		
	NSYSDNSAKATATKTOAEAKKFPDPSPIIQEAFOMVIOAEKDLKNTKPADGSDV		
	PQETTVGGSKQSGSSIGTSIRVSMLLDADENETASILMSCGFROMIHPNTENPDQAA		
	G"GGTAAQAAAKAAGDSDSAAALADALAAALGKAGQGGGILNALGQTASAAYVSA		
BASE COUNT	452 a	308 c	365 t
ORIGIN			
Query Match	97.08;	Score 1347;	DB 6; Length 1456;
Best Local Similarity	99.66;	Pred. No. 3.3e-313;	
Matches 1350; Conservative	0;	Mismatches 5;	Indels 0; Gaps 0;
QY	1	atgggttaatcctattggtccaggctctatagacgaacagacgacacacctccgcagat	60



Db 101 ATGGTTAACTCTATTGGTCCAGGTCCTATAGACGAACAAAGACGACACACCTCCCGCAGAT 160  
Qy 61 cttctctcaagattgagcgagtgacgaataaagatgagcgagctcaagaata 120  
Db 161 CTTTCTGCTCAAGGATTGGAGCGGAGTGCACAAATAAGATGCGGAAGCTCAAGAATA 220  
Qy 121 gcaggtgcggaagcctaagcctaaagaattctaaagaccgattctgtagagcgatgagcattc 180  
Db 221 GCAGGTGCGGAAGCTAAGCCTAAAGATCTAAGACCGATTCTGTAGACGATGGAGCATC 280  
Qy 181 ttgaggtctcagtggaatgctctcatgagctgtgagcagaataagctgggtattgctcttagt 240  
Db 281 TTGCGTCTGCGTGAATGCTCTCATGAGTCTGCAGATAAGCTGGGTATTGCTTCTAGT 340  
Qy 241 aacagctctcttactactagcagatctgcagacgtggaactcaagcagacgacgcacct 300  
Db 341 AACAGCTCGTCTTCTACTAGCAGATCTGCAGACGTGGACTCAACGACAGCGACCGCAGCT 400  
Qy 301 acgctctctccaccacgaagctctgattataagactcaagcgcaaacagcgttactgatact 360  
Db 401 ACGGCTCTCTCCACCACGCTTTGATGATTATAAGACTCAAGCGCAACACAGCTTACGATACT 460  
Qy 361 atctttacctcaacatacactagctgacatacagcgtctgttggtagcgtccagatgct 420  
Db 461 ATCTTTACCTCAACATCACTACTAGCTGACATACAGGCTGCTTTGGTAGGCTCCAGGATGCT 520  
Qy 421 gtcactaataaaaggaacacgagctactgaggaacacgcaatcgtcgagagtg 480  
Db 521 GTCATAATAAAGGATACAGCGCTACTGATGAGGAACCGCAATCGCTGCGGAGTGG 580  
Qy 481 gaaactaagaatgcgatgcaattaaagtgtgagcgcaaaattacagaattagcgaatat 540  
Db 581 GAAACTAAGAATGCGGATGCAATTAAAGTTGGCGCGCAAAATACAGAATTAGCGAAATAT 640  
Qy 541 gcttcgataaccagcgaattctgactcttttagttaaactgactctctcgacctta 600  
Db 641 GCTTCGGNATACCAAGCGATTTCTTGACTCTTTAGTAAACTGACTCTCTTCGACCTCTTA 700  
Qy 601 cagactgctcttcccaatctgtagcaacaataaacaagcagctgagcttcttaagag 660  
Db 701 CAGGCTGCTCTTCTCCAACTGTAGCAACAATAAACAAGCAGCTGAGCTTCTTAAAGAG 760  
Qy 661 atgcaagataaccagtagtccccagggaaacgcctgcaattgtctcaattctttagtgat 720  
Db 761 ATGCAAGATAACCCAGTAGTCCCAGGAAAACGCTTGAATTTGCTCAATCTTTAGTTGAT 820  
Qy 721 cagacagatgctacagcagacacagatagagaaagatggaatgcgattggagtgcatat 780  
Db 821 CAGACAGATGCTACAGCGACACACATAGAGAAAGATGGAATGGCATTAGGGATGCATAT 880  
Qy 781 ttgcagggacagcgttagtggagctgtgagaaatgctaaatctaaataacagataaagc 840  
Db 881 TTTCAGGACAGAACGCTAGTGGAGCTGTAGAAAATGCTAAATCTTAATAACAGATATAAGC 940  
Qy 841 aacataattcagctaaagcagcaatcgctactgctgaagacaaataagctgaagctcag 900  
Db 941 AACATAGATTGAGTAAAGCAGCAATCGCTACTGTAGACACAAATAGCTGAAGCTCAG 1000  
Qy 901 aaaaagttcccgactctcccaattctcaagaagcggaacaaatggaataacagcgtgag 960  
Db 1001 AAAAAAGTTCCCGACTCTCCAAATCTTCAAGACGCGGAACAATAAGGTAATACAGCTGAG 1060  
Qy 961 aagaatttcaaaataataaacctgcagatgggtctgagttcccaaatccaggaaactaca 1020  
Db 1061 AAAAGATCTTAAAAATATCAAACTGCAGATGGTTCTGATGTTCCTCAATCCAGGAATACA 1120  
Qy 1021 gttgaggtctccagcaagcaggaagtagtattgtagtattcgtgtttccatctgttta 1080  
Db 1121 GTTGGAGCTCCAAGCAACAGGAAGTAGTATTGGTAGTATTCGTGTTTCCATGCTGTGTA 1180  
Qy 1081 gatgatgtaaaatgagaccgcttccattttgatgtctgtggtttcgtcagatgattcac 1140  
Db 1181 GATGATGCTGAATGAGACCGCTTCCATTTTGATGCTGGGTTTCGTCAAGATGATCAC 1240

Qy 1141 atcttcaatacggaaaatcctgattctcaagctgcccaacagagctcagacacaagct 1200  
Db 1241 ATGTTCAATACGGAANAATCCTGATTCTCAAGCTGCCAACAGAGCTCGACACAAGCT 1300  
Qy 1201 agagcagcgaagccgctggagatgacagctgctgcagcgtgcagctgcagatgctcagaaa 1260  
Db 1301 AGAGCAGCAAGCCGCTGGAGTGCAGATGCTGCTGCAGCGCTGCAGATGCTCAGAAA 1360  
Qy 1261 gctttagaagcggctcttaggttaaagctgggcaacaacagggcatactactcaatgctttg99a 1320  
Db 1361 GCTTTAGAAGCGGCTCTAGTAAAGCTGGCAACAAACAGGCGATCTCAATGCTTTAGGA 1420  
Qy 1321 cagatcgcttctgctgctgttggtagcgcagaggt 1355  
Db 1421 CAGATCGCTTCTGCTGCTGTTGTGAGCGCAGGAGT 1455  
RESULT 8  
AX045133  
LOCUS AX045133 1852 bp DNA PAT 24-NOV-2000  
DEFINITION Sequence 3 from Patent WO0066739.  
ACCESSION AX045133  
VERSION AX045133.1 GI:11343732  
KEYWORDS  
SOURCE Chlamydomophila pneumoniae.  
ORGANISM Chlamydomophila pneumoniae  
REFERENCE 1 (bases 1 to 1852)  
AUTHORS Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.  
TITLE i(chlamydia) antigens and corresponding dna fragments and uses thereof  
JOURNAL Patent: WO 0066739-A 3 09-NOV-2000;  
Aventis Pasteur Limited (CA)  
FEATURES  
source 1..1852  
/organism="Chlamydomophila pneumoniae"  
/db\_xref="taxon:83558"  
CDS 1..1752  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC17378.1"  
/db\_xref="GI:11343732"  
/translation="MSLADKLGIASSNSSSSRSADVSTTATATPTPPPTFDYKT  
QAOTAYDTIFTSTSLADIAALVSLQDAVTNIKDAAATDEETAATAAEWETKNADVKY  
GAQITELAKYASDNQAILDSLGLTSDLLQALLOSVAANNKAAELLKENQDNVVP  
GKTPAIQSLVDQTDATATQIEKGNARIDRIFAGQNASGAVENAKSNNSISNDSAK  
AAATATQIAEAKQKFPDPSPILOEAEOMVIOAEKDLKNIKPADGSDVPNGFTTVGS  
KQGGSSIGSIRVSMLLDAAENETASILMSGFROMIHMFNTENPDQSAQAQOELAAQARA  
AKAAGDDSAALADAQALEALAKGQQGILNALGQIASAAVAVAGVPPAAASSI  
GSSVKQLYKTSKTSQDYKTQISAGYDAYKSI NDYGRANDATRDVNNVSTPALTR  
SVPRARTEARGPEKTDQALARVISGNSRTIGDVTQVSOYALQSVQITQSNPQANNEI  
ROKLTSAVTKPPQPGYPYVQLSNDSTOKFTAKLESFAEGSRTAAEIKALSFETNSLF  
IQOVLVNIIGSLYGLQ"  
BASE COUNT 578 a 409 c 407 g 458 t  
ORIGIN  
Query Match 83.2%; Score 1155.4; DB 6; Length 1852;  
Best Local Similarity 99.4%; Pred. No. 3.5e-267;  
Matches 1170; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
Qy 205 atgagctctggcagataaagctgggtattgcttcttagtaacagctgctcttactactagcaga 264  
Db 1 ATGAGCTGCGCAGATAAGCTGGGTATTGCTTCTAGTAACAGCTCGCTCTTACTACTAGCAGA 60  
Qy 265 tctcagacgtgtagctcaacgacgacgacccactacgctcctcccccacgtctgat 324  
Db 61 TCTCAGACGCTGGAGCTCAACGACGACGACCGCCTAGCGCTCCCTCCACCCACGCTTGAT 120  
Qy 325 gattataagactcaagcgcaaacagcttacgatactactctttacatcaacatcactagct 384



Db 121 GATTATAAGACTAAGCGCAACACGCTTACGATACCTATCTTTTACCTCAACATCACTAGCT 180  
Qy 385 gacatacaggctgttttggtgagcctccagatctgtcactaataataaaggatcacagc 444  
Db 181 GACATACAGGCTGCTTTGGTGAGCCTCCAGATGCTGTCTACTAATATATAAGGATACAGCG 240  
Qy 445 gctactgatgaggaacccgcaatcgctgagtgagggaacactaagaatgcccgtgcaatt 504  
Db 241 GCTACTGATGAGGAACCGCAATCGCTCGGAGTGGGAACATAAGATGCCGATGACGTT 300  
Qy 505 aaagtggcgcgcgaataatcacgaattagcgaatatgtcttggtaataccagcattctt 564  
Db 301 AAAGTTGGCGCGCAATTACGAATAGCAAAATAGCAAAATAGCTTCGGATAACCAAGCGATTCTT 360  
Qy 565 gactcttagttaacagctactcttcgacctttacagactgtcttctccaatctgta 624  
Db 361 GACTCTTTAGGTAACAGTACTCTTCGACCTCTTACAGGCTGCTTCTCTCAATCTGTA 420  
Qy 625 gcaacaataaacaagcagctgagcttcttaaaagatgcaagataacccagtagtccca 684  
Db 421 CCAACATATAACAAGCAGCTGAGCTTCTTAAAGAGATGCAAGATAACCCAGTAGTCCCA 480  
Qy 685 gggaaaaacgctgcaattgctcaatcttttagttgatcagacagatgctacagcgacacag 744  
Db 481 GGGAAAACGCTGCAATTCCTCAATCTTTAGTTGATCAGACAGATGCTACAGCGACACAG 540  
Qy 745 atagagaagatgaatcgatgggagtcataatttttcagacagacagcgttagtga 804  
Db 541 ATAGAGAAGATGGAATGCGGATAGGATGCAATTTTGGCAGGACAGACGCTAGTGGG 600  
Qy 805 gctgtagaaaatgctaaatctaaataacagataaagcaacatagattcagctaaagcagca 864  
Db 601 GCTGTAGAAAATGCTAAATCTAATAACAGTATAAGCAACATAGATTACAGTAAAGCAGCA 660  
Qy 865 atcgtactgtagagacaaaatagctgaagctcagaaaagttcccgactctccaatt 924  
Db 661 ATCGTACTGTAGACACAAATAGCTGAAGCTCAGAAAAGTTCCCGACTCTCCAAAT 720  
Qy 925 ctcaagaagcgaacaatgtaatacagctgagaaagatcttaaaatcaaacct 984  
Db 721 CTTCAAGAAGCGGAACAATGGTATACAGGCTGAGAAGATCTTAAAAATATCAAACT 780  
Qy 985 gcagatggtctgtatgttccaaatccaggaactacagttgaggtcccaagcaacaagga 1044  
Db 781 CGAGATGTTCTGATGTTCCAAATCCAGGAACACTACAGTTGGAGGCTCAAGCAACAAGGA 840  
Qy 1045 agtagtattgtagtattcgtgtttccatgctgtttagatgctgtaaaatgagaccgt 1104  
Db 841 AGTAGTATTGGTAGTATTCTGTTTCCATGCTGTGTAGATGATGCTGAAAATGAGACCGCT 900  
Qy 1105 tccattttagtctgggttcgtagatgattccatgtttcaatcacggaataatcctgat 1164  
Db 901 TCCATTTTGTGTCGTGGGTTTCGTGAGATGATTCACATGTTCAATACCGGAATTCCTGAT 960  
Qy 1165 tctcaagctgcccaacagagctgcagcagcaaaagctagagcagcaaaagcgcgtggagat 1224  
Db 961 TCTCAAGCTGCCCAACAGAGCTCCAGCAACAAGCTAGAGCAGCGAAAGCGCTGGAGAT 1020  
Qy 1225 gacagtgtcgtgcagcgtgagcagatgctcagaaaagcttttagaagcggctctaggtaaa 1284  
Db 1021 GACAGTGTGTCAGCGCTGSCAGATGCTCAGAAAAGTTTAGAAGCGGCTCTAGGTAAA 1080  
Qy 1285 gctgggcaacaacagggcatactaatgctttggacagatcgcttctgctgtgttg 1344  
Db 1081 GCTGGGCAACACAGGGCATACTCAATGCTTTAGGACAGATCGCTTCTGCTGCTGTGTG 1140  
Qy 1345 agcgcaggag-tccttcccgctgcagcaagttctatgg 1380  
Db 1141 AGCGCAGGAGTTCTCTCCCGCTGCAGCAAGTCTATAG 1177

RESULT 9  
AR144067

LOCUS AR144067 150 bp DNA PAT 08-AUG-2001  
DEFINITION Sequence 8 from patent US 6210876.  
ACCESSION AR144067  
VERSION AR144067.1 GI:15105934  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 150)  
AUTHORS Cerney, M.B.  
TITLE Nucleic acid primers and probes for detecting Chlamydia pneumoniae  
JOURNAL Patent: US 6210876-A 8 03-APR-2001;  
FEATURES Location/Qualifiers  
1..150  
source /organism="unknown"  
BASE COUNT 43 a 34 c 40 g 33 t  
ORIGIN  
Query Match 10.6%; Score 146.8; DB 6; Length 150;  
Best Local Similarity 98.7%; Pred. No. 1.1e-24;  
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 366 tacctcaacatcactagctgacatcacaggctgtttgtgtgagcctccaggatgctgtcac 425  
Db 1 TACCTCAACATCACTAGCTGACATACAGGCTGCTTTGGTGAGCCTCCAGGATGCTGTGAC 60  
Qy 426 taatataaagatcacagcgctactgatgagaaacacgcaatcgctgcgagtggaac 485  
Db 61 TAAATATAAGGATACAGCGGCTACTGTAGTGAAGAACCCCAATCGCTGGGTGTGGGAAC 120  
Qy 486 taagaatgccgatgcaattaaagtggcgc 515  
Db 121 TAAGATGCCGATGCAGTTAAGTTGGCGC 150  
RESULT 10  
AE001333/c  
LOCUS AE001333 10954 bp DNA BCT 30-OCT-2000  
DEFINITION Chlamydia trachomatis section 60 of 87 of the complete genome.  
ACCESSION AE001333 AE001273  
VERSION AE001333.1 GI:3329068  
KEYWORDS  
SOURCE Chlamydia trachomatis.  
ORGANISM Chlamydia trachomatis  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE 1 (bases 1 to 10954)  
AUTHORS Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,  
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
Koonin, E.V. and Davis, R.W.  
TITLE Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis  
JOURNAL Science 282 (5389), 754-759 (1998)  
MEDLINE 99000809  
PUBMED 9784136  
REFERENCE 2 (bases 1 to 10954)  
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,  
Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE 3 (bases 1 to 10954)  
AUTHORS Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,  
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
Koonin, E.V. and Davis, R.W.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University  
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA  
FEATURES Location/Qualifiers  
1..10954  
/organism="Chlamydia trachomatis"  
/strain="D/UW-3/CX"



```

/gene
complement(146. .2089)
/gene="CT622"
complement(146. .2089)
/gene="CT622"
/codon_start=1
/transl_table=11
/product="CHLPN 76kDa Homolog"
/db_xref="GI:3329069"
/translation="MESGPESVSSNOSMNPPIINGQIASNETKSTKESESPASS
SVSFTSLAKRALISLRDAIINKNSPTDSLSQLEASTSTSTVTTRVAARDNEAKS
NFTAKSGNQTLLAETKMDALQDMERLAKQAEVTTRKEALQKEQVEIDK
LNLQVLEKQNTLKETLTITTSADQITPAINSOLEINKNSADQIILKDEGONISYEA
LTVAGVEIKASSAGIKLGOALQSIIVDAQSOAAVLQAQONNSPDNIATKLIDAA
ETVKNELQHTGLTDSPLVKAEQISOAKQDIOETKPSGSDIPIVGPSASAAGS
ACALKSSNNGSRISILLDDVDNEMRAIAMQGRSMIEQFNVNPNATAKLOAMEAQL
TAMSDQLVAGDELPAETQAIKALQAQLKPSDTGLATAMQVAFAAKLVGGSGAGT
AGTVQNVKOLYATFSSSTSSSYAALSQGYSAIKTNSLYSESRSYGVSQASISQAN
PALRSYSRSGIESQGSADASQRAAETIVRDSOTLGDVYSRLQVLDLSMTISVNPQ
VNOEETMQKLTAISIKAPQFGYPAVONSADSLQFAAQLEREFVDSRSLAESRENAF
RKQPAFTQOVLVNIASLFCGLS"
complement(2197. .3495)
/gene="CT623"
complement(2197. .3495)
/gene="CT623"
/codon_start=1
/transl_table=11
/product="CHLPN 76kDa Homolog"
/db_xref="GI:6578109"
/translation="MKKFIYKFGVGLLACAGTSLNLAFAQAASSMDSQSVEDLDSYL
SSGFVEYTRKRDVLAGDVRAWYAKEDLETTOTPAKPMPLTNRYSRSEFNLVDY
TAANSWTKMKNVTIAGGESSAAGLDINRAFLGYRFYKNPETOAEVFAEIRNGLGD
IPDSVQFNSNFDGILHYAARRISEKLPFTMIYHGGPFVNVMAEKVAVWVEAILNKL
PGNFVKTSVVDNHTLTAKTNDPADASAAQAPAKPKYDLYLWQNLVCKSTAMPWFNG
QTKNLTYGAYLNPDAEPENKQSTPTTKITNGKENHAWFICGSLGGVRRAGDWS
ATVRYEVEALAIPEIDVAGIGRGNQMYFAQAIKQGLDPKESNGFTNYKGVSYQFV
MGUTDSVFRAYAAYSKPANDNLGSDFTYRKYLGLISF"
complement(3738. .5348)
/gene="mvln"
/notes="CT624"
complement(3738. .5348)
/gene="mvln"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/db_xref="GI:3329071"
/translation="MRKDDGSLVRSFLNLLSGTFFSRFLGMLREIVMATYFGADPLV
ASFLAFRTIFFRLKLGPIGLGAFIPHEFLRAQNISRAVFFRSFRFFCYSAIL
FTLIIELGLCWCSCVTGSLFDLTLLTIIILPSGIFLMYTVNSTLHCEKFKFSVGL
APSVNVNWTGTVFLARNYDPRNRIFGLAVLVIGFLEWAVTLPGVMKFLGOSKEVP
QERDSTRALAPLSGLISMGIFQLNLLCDMLARYINEVGPLYLMYSVRIQQLPVHL
FGLGVTLPLPALSRCVQDQHQGDLLRFLSKLIVAVNVMTWGLLLPALPGVRVL
YENGVPFKTAHAIVEIRGYSIIPMALAPLVSAFVARRNYKVPMLVGIIAAVVN
MVLNVJGLCVKQVAYLATVSLVSMGQALMYCAGKSPTYKGLMRTFKESGKTV
ITVILAAVTIGVNIHTTVYVIEPLTVPTKPLVSLFSDQCGVFAESAFLFLSVLFG
LAKLLKEDLMLNLISFYWKGHQSILRN"
5521. .6387
/gene="nfo"
/notes="CT625"
5521. .6387
/gene="nfo"
/codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AAC68229.1"
/db_xref="GI:3329072"
/translation="MFLPPQFALLGAHTSAAGLHNLALYEGRIDGATTVOLFTANQ
ROMKRRTLTOEWDPQAFIALNETSLVIMSHAGYLNPNPAPNPEILKTKVCMHQAETA
DCISLGISFVNHPHGAALSDSKESCLDRITASFQMAPLFENNPNPLVLETTAGOGS
LIGSSFEELAYLQGIKAHPIGVCLDTCIFPAAGYDISSVAGVEQVLKHFDAVIGLS

FLRAIHLNDSVFALGKNKRHAPIGECCIGSDSFCFLMQDQTRMLPKYLETPGGPDL
WTKEIRYLQKVC"
complement(6484. .7113)
/gene="rs4"
/notes="CT626"
complement(6484. .7113)
/gene="rs4"
/codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AAC68230.1"
/db_xref="GI:3329073"
/translation="MARYCGPKNRIARRFGANIFGRGNPLLRPNPPGQHQMKRKKK
SDVGLQEEOKLCAKGMILEKOLVAYKEVNVKNQNAOMLEKTECRLDNIVYRL
GFAKTIFAAQOLKSHGVHLVNGKVDRRSPFVRPMOISLKEKSKRLAIVTESLENKD
OSSLPAYLSLIDKAAFKGELVVAPELDQIASQLPLPVNVSVICEFLSHRT"
7497. .8480
/gene="yceA"
/notes="CT627"
7497. .8480
/gene="yceA"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC68231.1"
/db_xref="GI:3329074"
/translation="MEKNYVALAYVYFVPSNPHEETALHKOLFCTMDVSCRIVISEE
GNGQESGYOPDAERYMAWLKQRFDAISIKFIHIEENTFPRTVYKRELVALGCS
VDTTKQKHISPEWHEKRLQENRCLVDVRNNYEMKIGHFENAVLDTETFPEDPA
DLRAKHDPARTPVMYCTGGICRELYSALLKEGFEVQLDGGVYAYGLKMTGW
RKLFFVDDRRAMPIDEADPNVSIARCSLCNTSDTYYNCACTDNNLFTCCESCIA
THKGCSSECSQAPRIADFSAERGNKPKRRKHLCTTIEQSCCLKEQENQPA"
complement(8552. .9427)
/gene="ispa"
/notes="CT628"
complement(8552. .9427)
/gene="ispa"
/codon_start=1
/transl_table=11
/product="Geranylgeranyl pyrophosphate synthase"
/protein_id="AAC68232.1"
/db_xref="GI:3329075"
/translation="MDYFSVTRSVKVKLRDLSLQDFQNTGGLRDPIEYALLGGKRV
RPLLVLGFAEGVHKRDLVDTAIEYIHTSTLIADLPCMDDDMRGRKPSVHKAFD
EASALLASYALIPAAAYARIKNKAKALKAVVSCQEDVEAEYEDILELIEFRGVGILG
GOYEDIFFHDFSKENVLGIITKKTGALFEIACVSGWLFGGGERESSTLVAFEAHFGL
LFOIRDDLADLSQDDQEKHMNYALLFEVAAKELLDSHESFCIKNLHLLQERGLECL
EPMELCKNVFCGWK"
complement(9483. .10100)
/gene="glmU"
/notes="CT629"
complement(9483. .10100)
/gene="glmU"
/codon_start=1
/transl_table=11
/product="UDP-GlcNAc Pyrophosphorylase"
/protein_id="AAC68233.1"
/db_xref="GI:3329076"
/translation="MVLSSLSFSEPEELYPEIVSOAEFVMSILLLEKLAHPTFSGI
HGHLBEGVYLKKNKEITIEQGAAYVESGAYICGPCIIGYTVQVRHGAIRGVYITSSHC
VHGHCSEIKNSYLGHHAKAAHFAVYGVSDSVNGLAGVRCANFRDLGKRTIFFHHS
ERCNTKRKILGAFGLRGVSGVNCNTVLPNPGCYVASATKILPNQTIY"
complement(10153. .10836)
/gene="cpxR"
/notes="CT630"
complement(10153. .10836)
/gene="cpxR"
/codon_start=1
/transl_table=11
/product="HTH Transcriptional Regulatory Protein and
Receiver Domain"
/protein_id="AAC68234.1"
/db_xref="GI:3329077"

```











Search completed: February 7, 2002, 19:33:50  
Job time: 13544 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:48:06 ; Search time 12230.8 Seconds  
(without alignments)  
3432.748 Million cell updates/sec

Title: us-09-391-606-1

Perfect score: 2545

Sequence: 1 ttgcggtgtgttaacggtg.....tgccactccactgtctcttt 2545

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.om.\*

20: em.or.\*

21: em.ov.\*

22: em.pat.\*

23: em.ph.\*

24: em.pl.\*

25: em.ro.\*

26: em.sts.\*

27: em.sy.\*

28: em.un.\*

29: em.vi.\*

30: em.htgo.hum.\*

31: em.htgo.inv.\*

32: em.htgo.rod.\*

33: em.htg.hum.\*

34: em.htg.inv.\*

35: em.htg.rod.\*

36: em.htg.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No	Score	Match	Length	DB	ID	Description
1	2225.2	87.4	2238	6	AX045137	AX045137 Sequence
2	2122.2	83.4	10421	1	AE001654	AE001654 Chlamydia
3	2122.2	83.4	12548	1	AE002165	AE002165 Chlamydia
4	2122.2	83.4	300550	1	AP002547	AP002547 Chlamydia
5	2109.2	82.9	2148	1	CHT76KDA	L23921 Chlamydia p
6	1459.4	57.3	2156	6	AX045131	AX045131 Sequence
7	1447	56.9	1456	6	AX045135	AX045135 Sequence
8	1155.4	45.4	1852	6	AX045133	AX045133 Sequence
9	264	10.4	14698	1	AE002357	AE002357 Chlamydia
10	257	10.1	10954	1	AE001333	AE001333 Chlamydia
11	217.6	8.6	3455	1	CTU50732	U50732 Chlamydia t
12	200.2	7.9	6849	12	AY029367	AY029367 Synthetic
13	171	6.7	5108	6	AX138932	AX138932 Sequence
14	171	6.7	5108	6	AX138933	AX138933 Sequence
15	171	6.7	5111	6	AX138936	AX138936 Sequence
16	171	6.7	5185	6	AX138928	AX138928 Sequence
17	171	6.7	5188	6	AX138940	AX138940 Sequence
18	171	6.7	5254	6	AX138941	AX138941 Sequence
19	149.2	5.9	4928	6	AR141727	AR141727 Sequence
20	149.2	5.9	4928	6	I49890	I49890 Sequence 1
21	146.8	5.8	150	6	AR144067	AR144067 Sequence
22	145	5.7	7475	6	AX001323	AX001323 Sequence
23	145	5.7	7475	6	E36259	E36259 Human semap
24	145	5.7	8192	6	AX001324	AX001324 Sequence
25	145	5.7	8192	6	E36260	E36260 Human semap
26	142	5.6	3125	6	AR034783	AR034783 Sequence
27	142	5.6	7108	6	AX001326	AX001326 Sequence
28	142	5.6	7108	6	E36262	E36262 Human semap
29	139.8	5.5	4276	6	AX154496	AX154496 Sequence
30	139.8	5.5	4282	6	AX164583	AX164583 Sequence
31	139.8	5.5	4328	6	A92666	A92666 Sequence 2
32	139.8	5.5	4328	6	AR137192	AR137192 Sequence
33	139.8	5.5	4818	6	A92668	A92668 Sequence 4
34	139.8	5.5	5107	6	A92667	A92667 Sequence 3
35	139.8	5.5	9600	6	A92665	A92665 Sequence 1
36	139.2	5.5	5459	6	AX154499	AX154499 Sequence
37	139.2	5.5	5882	6	AX154501	AX154501 Sequence
38	139.2	5.5	7731	6	A94048	A94048 Sequence 29
39	139.2	5.5	7731	6	A94061	A94061 Sequence 42
40	139.2	5.5	7731	6	AX011133	AX011133 Sequence
41	139.2	5.5	7731	6	AX011146	AX011146 Sequence
42	139.2	5.5	8068	6	A94046	A94046 Sequence 27
43	139.2	5.5	8068	6	A94054	A94054 Sequence 35
44	139.2	5.5	8068	6	AX011131	AX011131 Sequence
45	139.2	5.5	8068	6	AX011139	AX011139 Sequence

#### ALIGNMENTS

#### RESULT 1

AX045137

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

PAT

DNA

Sequence 7 from Patent WO0066739.

AX045137

AX045137.1 GI:11343736

Chlamydia pneumoniae.

Chlamydia pneumoniae

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

1 (bases 1 to 2238)

Murdin, A.D., Oomen, R.P., Wang, J. and Dunn, P.

-(Chlamydia) antigens and corresponding dna fragments and uses thereof

Patent: WO 0066739-A 7 09-NOV-2000;

Aventis Pasteur Limited (CA)

Location/Qualifiers

1..2238

/organism="Chlamydia pneumoniae"

/db\_xref="taxon:83558"



```

CDS      766..2238
          /note="unnamed protein product"
          /codon_start=1
          /transl_table=11
          /protein_id="CAC17330.1"
          /db_xref="GI:11343737"
          /translation="MVNPIGPGPIDETERTPPADLSAQGLEASAAKSAQAQRIAGAE
          AKPKESKTSVVERWSILRSVAVNALMSLADKLGIASSNSSSTSRSDVSTTATATPP
          PPPTFDYKTKQAOAYDTIFSTSLADIQALVSILOPVTNIKDVTATDETAIAEWM
          ETKNADVKVQAQITELAKYASDNQAILDSGLKLTSPDLLQALLQSVANNKRAEUL
          KEMDNVPVKGPTAIQSLVDQDATQIEKDNINAIQDAYFAGNAGAVENAKNS
          NSINIDSSAKQIAIAKTQIAEKFPDPSPIQEAQOMVIOAEKDLNKPADGSDV
          PNPQTTVGGKQKQSDSISIRVSMLLDQENETASILMSGFROMIHFNFTENPDQAA
          QOELAAQAKAAGDQDAAAALADOKALEALGKAGQOQILNALGQIASAAVSA
          GVLPQQVLIRARQYAVVEQKLISEEDLNSAVDHHHHH"
BASE COUNT 689 a 459 c 497 g 593 t
ORIGIN

Query Match      87.4%; Score 2225.2; DB 6; Length 2238;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2230; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 175 atgacaaaaacattatgcttgggtgttagaaggattctcaatcggtttgcctaaacag 234
Db 1 ATGACAAAAACATATGCTGGGTGTAGAGGGATTCTCAATCGTTTGCCTAAACAG 60

QY 235 ttttttgtaaatgtagtgtgtgcagtggaacacattcggttcttcagaaacctccact 294
Db 1 TTTTGTGTAATGTAGTGTGTCGACTGGACACATTCGTTCCTTCAGAACCTCCACT 120

QY 295 acagaaaaagctgtacaaaagctatgaatacaaaactgtgttggcagtgctcgtc 354
Db 1 ACAGAAAAAGCTGCTACAAACGCTATGAATACAAATCTGTTGGCAGTGGCTCGTC 180

QY 355 gaaagacatagtcaggttcttggatcaatggacagaaaagcctctatatctttatgga 414
Db 181 GGAAGACATAGTCAGGTTCTTGGATCAATGACAGAGAAAAGCCTCTATATCTTTATGGA 240

QY 415 gcttcttaataaacctttagcaagctcagaaagctacgttaaatggaaaaagaaac 474
Db 241 GCTTTCTTAATGAACCTTTAGCAAGGCTACGAAAGACTAGCTTAAATGGAAGAAAAC 300

QY 475 ctagcttggtttattggagaaactttaggggactcagaaaagctggagactggctgccc 534
Db 301 CTAGCTTGTTATTTGGAGAACTTTAGGGGACTCAGAAAAGCTGGAGACTGGTCTGCC 360

QY 535 acagtcagttatgtagtgcgaagccttgcagttccagaaaatagatgttttcagggatt 594
Db 361 ACAGTACGTTATGATGTGCAAGCCTTGTGGGTTCCAGAAATAGATGTTTCAGGGATT 420

QY 595 ggcctgtgtaatttattaaagtttgggtgcccgaagaactgtgctaaactatgatcct 654
Db 421 GGCCTGTGTAATTTTAAAGTTTGGTTCGCCCAAGCAATTCGTCTTAACATGATCCT 480

QY 655 aaagaggcctaagttttacaaaattataaaggattttccgctctatatatatgtggcgc 714
Db 481 AAAGAGGCTAATGGTTTACAAATATAAAGGATTTCCGCTCTATATGATGGCAATC 540

QY 715 acagattctctatcattcagagcttatgggcttactccaaaccagcaaaacgataaacctc 774
Db 541 ACAGATTCTATCATTCAGAGCTTATGGGCTTACTCCAAACACAGCAACGATAACTC 600

QY 775 ggcaggtatttacttccgaagtttgatctagtgataaattcagcgttttaagtcaaa 834
Db 601 GGCAGTGATTTTACTTTCGAAAGTTTGAICTAGTGATATAAATTCAGCGCTTTAAAGTCAA 660

QY 835 ttttaataaaactcttaaaacagctcgcatattattagtgagagctttttttat 894
Db 661 TTTTAATAAAATCTTTAAACACAGCTCGCATTAATTAATAGTGAGAGCTTTTTTTTTAT 720

QY 895 tttttataaaaaactaaaagattttatttttttgagtttttttttatggttaactcatt 954

```

```

Db 721 TTTTATAATAAACTAAAGATTTTATTATTTTGGTGTATGTTTAACTCATT 780
QY 955 ggtccaggtcctatagacgaaacagacgacacactcccgagatcttcttgcctcaagga 1014
Db 781 GGTCCAGGCTCTATAGACGAAACAGACGACACTCCCGCAGATCTTCTTCGCTCAAGGA 840
QY 1015 ttgagggcagtgagcaaaaataagagtcggaagctcaaaagataagcaggtgcggaagct 1074
Db 841 TTGGAGGCGAGTCAGCAAAATAAGAGTCGGAAGCTCAAGAATAAGCAGGTGCGGAAGCT 900
QY 1075 aagcctaaagaatctaagaccgattctgtagagcagtgagcattcttgccttgcagtg 1134
Db 901 AAGCTAAAGAATCTAAGACCGGATTCTGTAGAGCGATGAGCATCTTCGCTTCGCACTG 960
QY 1135 aatgctctcatgagcttgcaagataaagctgggtattgcttcttagtaaacagctcgtctct 1194
Db 961 AATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGTAACAGCTCGCTTCT 1020
QY 1195 actagcagatctgcagacgttgactcaacgacagcgcacacctacgcttctctccccc 1254
Db 1021 ACTAGCAGATCTGCAGAGTCGACTCAACGACAGCGACCGCACTACGCTTCTCCACCC 1080
QY 1255 acgtctgattataagactcaagcgcgaacagcttacgatactattcttacotcaaca 1314
Db 1081 ACGTTTGATGATTATAAGACTCAAGCGCAACAGCTTACGATACCTATCTTTACCTCAACA 1140
QY 1315 tcaactagctgacatacagagctgcttggtagcctccaggtgctgtcactaataaag 1374
Db 1141 TCACCTAGCTGACATACAGGCTGCTTGGTAGCCTCCAGGATGCTGCTCACTAATAAAG 1200
QY 1375 gatacagcgctactgatgagaaaccccaatcgctcgaggtgggaactcaagaatgcc 1434
Db 1201 GATCAGCGGCTACTGATGAGAAACCCCAATCGCTGCGGAGTGGGAACCTAAGAAATGCC 1260
QY 1435 gatcaattaaagttggcgcaaatattacagaattagcgaatatgcttcggataaccaa 1494
Db 1261 GATGCACTTAAAGTTGGCGCAAAATTACAAATTTAGCGAAATATGCTTCGATAAACCA 1320
QY 1495 gcgattctgactctttaggttaactgactccttcgacctcttacagactgctctctc 1554
Db 1321 GCGATTCCTGACTCTTTAGGTAACCTGACTTCTTCCGACCTCTTACAGGCTGCTCTCTC 1380
QY 1555 caactctgtagcaacaataacaaagcagctgagctctttaaagagatcaagaataaccca 1614
Db 1381 CAATCTGTAGCAACAATAACAAAGCAGCTGAGCTTCTTAAAGAGATGCAAGATAACCCA 1440
QY 1615 gtagtccagggaaaaacgctgcgaattgctcaatttttagttgatcacagatgctaca 1674
Db 1441 GTAGTCCCGAGGAAAAACGCTGCAATTGCTCAATCTTTAGTTGATCAGACAGATGCTACA 1500
QY 1675 gcgacacagatagagaaagatggaaatgcgattggggatgcatatttttgcaggacagaa 1734
Db 1501 GCGACACAGATAGAGAAAGATGGAAATCGGATTAGGGATGCTATATTTTGCAGGACAGAAC 1560
QY 1735 gctagtgagctgtagaaaaatgctaaatctaaacagataaagcaacatagattcagct 1794
Db 1561 GCTAGTGAGCTGTAGAAAAATGCTAAATCTTAATACAGATTAAGCAACATAGATTCAGCT 1620
QY 1795 aaagcagcaatcgctactgctaagacacaaaatagctgaagctcagaaaaagttcccccag 1854
Db 1621 AAAGCAGCAATCGCTACTGCTTAAGACACAAATAGCTGAAGCTCAGAAAAAGTTCCCGAC 1680
QY 1855 tctccaatcttcaagaagcggaaacaaatgtaatacagagctgagaaagatctttaaataat 1914
Db 1681 TCTCCAAATCTTCAAGAAAGCGGAAACAAATGGTAAATACAGGCTGAGAAAGATCTTAAAT 1740
QY 1915 atcaaacctcgagatggtctgagttccaaatccagaaactcagttggaggtcccaag 1974
Db 1741 ATCAAACTGCGAGATGGTCTGATGTTCCAAATCCAGGAACCTACAGTTGGAGGCTCCAG 1800
QY 1975 caacaagaagtagtattggtagtattcgtgtttccatgctgttagatgtagtgaataat 2034
Db 1801 CAACAAGGAAGTAGTATTGGTAGTATTGCTGTTTCCATGCTTGTAGATGATGCTGATAAT 1860

```







```

/transl_table=11
/product="hypothetical protein"
/protein_id="AADI18870.1"
/db_xref="GI:4377036"
/translation="MSPLLIKPFSPPLKQDYLFDISPYTSSEITIGGSYFKLNKASL
QSSFLRURSTIIIS"
gene      8654..9535
/feature   /gene="nfo"
CDS       8654..9535
/feature   /note="CPn0732"
/feature   /gene="nfo"
/feature   /codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AADI18871.1"
/db_xref="GI:4377037"
/translation="MKVLPPPSIPLLAHTSTAGLKNALIVEGRDIGASTVQLFTANQ
RQWRRALKREVIEDFAALKKETDLSYIMSHAGYLINPGADPDVILKSKRIGTYOETL
DCITLIGTSFNHFPCAALKSKEDCMNKIYVSFSQAPLFDSPPLVLLTETAGQCT
LIGNFELHGLYVONLNQPIPIGVCDTCHIFAAGYDITSPOCHWEDVLEFDEVGLS
YLRAFHLDSMPFLGANKDRHAPLGEYIGKESFKLMTDERTKKIKPYLETGGPEN
WKEIGELLKFSKNRDS"
gene      complement(9654..10283)
/feature   /gene="rs4"
CDS       complement(9654..10283)
/feature   /note="CPn0733"
/feature   /gene="rs4"
/feature   /codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AADI18872.1"
/db_xref="GI:4377038"
/translation="MARVCGPKNVARREFGANIGRSNPLLLKPKHPGOGHOMRKKK
SDYGLIIEKOKLKACYMIMKQLVAFKEVIRKQGNVAMFLERECRLDNWYRM
GFAPKIFBAQQLVAHGHILVNGRRVDRSFFLRPMQISLKEKSRKQSVKDALESKD
RSSLSYSILDKTFKGLLVSPEDQTEAQLPLINISVYVCEFLSHRT"
BASE COUNT 3278 a 2158 c 1983 g 3002 t
ORIGIN

```

```

Query Match      83.4%; Score 2122.2; DB 1; Length 10421;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2138; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 174 catgacacacacattatgcttggtttagaaggattccaatcgcttccctaaaca 233
Db 5799 CATGACAAAAACATATGCTTGGGTTGTAGAAGGGATTCATATCGTTGCTTAACA 5740

QY 234 gttttttgaaatgtagtctgactggaacacattcgcttccctcagaacccctccac 293
Db 5739 GTTTTTGTGAATGTAGTGTGCTGACCTGGAACACATTCGTTCTTCAGAAACCTCCAC 5680

QY 294 tacagaaaaagctgctacaaacgctatgaaatacaataactgtgttgagtgagctcgt 353
Db 5679 TACAGAAAAAGCTGCTACAAACGCTATGAATAACAATACTGTGTTGGCAGTGGCTGT 5620

QY 354 cgaagacatagtcaggttcttggatcaatgacagaaaaagcctctatatctttatgg 413
Db 5619 CGNAAGACATGATCAGGTTCCTTGGATCAATGGACAGAAAAAGCCCTATATCTTATG 5560

QY 414 agctttctaatgaaccccttagcaaggctacgaagactacgtttaaataggaaaaa 473
Db 5559 AGCTTTCTTAATGAACCTTTAGCAAGGCTACGAAGACTACGTTAAATGGAAAAAGAAA 5500

QY 474 cctagctgtttattgagaaacttttagggactcagaagactggagactgggtctgc 533
Db 5499 CCTAGCTTGGTTATTGGAGGAACCTTTAGGGGACTCAGAAAAAGCTGGAGCTGGTCTGC 5440

QY 534 cacagtacgttatgtagtgcgaagccttgctcagttccagaaatagatgtttcagggat 593
Db 5439 CACAGTACGTTATGAGTATGCGAAGCCTTGTGCGTTCCAGAAATAGATGTTTCAGGGAT 5380

QY 594 tggccgtggttaatttataagtttgggttcgcccaagcaattgctgctaactatgatcc 653

```

```

Db 5379 TGGCCGTGGTAATTTATTAAAGTTTGGTTCGCCCAAGCAATTCGTCTAACTATATGCC 5320
QY 654 taaagagcctaagtagtttttcaaatataaaagattttccgctctatatatgatggcat 713
Db 5319 TAAAGAGCGCTAATGGTTTTTACAAATATAAAGGATTTTCGCTCTATATATGATGSCAT 5260
QY 714 cacagattctctatctcagagcttatgggcttactccaaccagcaaacatataaact 773
Db 5259 CACAGATTCTATCATTCAGCTTATGGGGCTTACTCCAAACCAGCAACCAATAACT 5200
QY 774 cggcagtgattttactttccgaaagtgtgatctaggtatataatttcacgcttttaagcaa 833
Db 5199 CGGCAGTGATTTTACTTTCCGAAAGTTTGATCTAGGTATAATTTACAGCGTTTTAAGTCAA 5140
QY 834 attttaataaaacttttaaaaaacaggctcgacatttaatttagtgagagctttttttta 893
Db 5139 ATTTTAATAAAATCTTTAAAAACAGGCTCGCATTAATTAATAGTGAGAGCTTTTTTTTTA 5080
QY 894 tttttataataaaactaaagatttttattattttttgagtttttttgggttaactctat 953
Db 5079 TTTTTTATAATAAACTAAAGATTTTATTTATTTTGTGAGTTTTTATGTTTATGTTAATCTAT 5020
QY 954 tggctcaggtcctatagacgaaacagacacacccctcccgacagatttttctgctcaagg 1013
Db 5019 TGGTCCAGTCTCTATAGACGAACAGACACACACCTCCCGCAGATCTTTCTGCTCAAGG 4960
QY 1014 attgagcgagtagcgacaaataaagtgaggagctcaagaatacgaagtcggaagc 1073
Db 4959 ATTGGAGCGAGTGCAGCAATAAAGAGTGGGAAAGCTCAAAAGATAGCAGGTGCGGAAGC 4900
QY 1074 taagcctaaagaacttaagaccgctctgttagagcagatggagcattcttggtctgcagt 1133
Db 4899 TAAGCCTAAAGAACTTAAGACCGATTCTGTAGAGCGATGGAGCATCTTGCCTTCTCAGT 4840
QY 1134 gaatgctctatgagtcgtgagataaagctgggtattgcttctagtaaacagctgcttc 1193
Db 4839 GAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTTASTAACAGCTGCTCTC 4780
QY 1194 tactagcagatctcagacgtggactcaacgacagcagcaccctacgcctcctccacc 1253
Db 4779 TACTAGCAGATCTGCAGACGTGGACTCAACGACAGCAGCAGCAGCAGCAGCAGCAGCAGC 4720
QY 1254 cagctctgatattataaagactcaagcgcaaacagcttaccatactatttcaactcaac 1313
Db 4719 CACGTTTTCATGATTATAAGACTCAAGCGCAAAACAGCTTACGATACATCTTTACCTCAAC 4660
QY 1314 atcaatagctgacatacaggctgctttagtgagcctccagagatgctgctcaataataaa 1373
Db 4659 ATCACTAGCTGACATACAGGCTGCTTTGGTGAGCCTCCAGGATGCTGCTCAATATAAAA 4600
QY 1374 ggatacagcggctactgatgagaaacccgcaatcgctcgagtgaggaaactaagaatgac 1433
Db 4599 GGATACAGCGGCTACTGATGAGAAACCGCAATCGCTCGGAGTGGGAAACTAAGAAATGC 4540
QY 1434 cgaatgaattaaagtggcgcgcaaatcacagaattagcgaataatgottcggataaaca 1493
Db 4539 CGATCAGTTAAAGTTGGCGCGCAAAATACAGAAATAGCGAAATATGTTTCGGATAACCA 4480
QY 1494 agcgattcttgactcttaggttaaactgaactccttcgacctctacagactgctcttct 1553
Db 4479 AGCGATTCTTGACTCTTTAGGTAAACTGACTTCTTCGACCTCTTACAGGCTGCTTCT 4420
QY 1554 ccaactctgtagcaaacataacaaagcagctgagcttctttaaagagatgcaagataaacc 1613
Db 4419 CCAACTCTGTAGCAAAACATAACAAAGCAGCTGAGCTGCTTCTTAAAGAGATGCAAGATAACC 4360
QY 1614 agtagtccagggaaaaacgctgcaattgctcaactcttagttgacacagatgctac 1673
Db 4359 AGTAGTCCAGGGAAAAACGCCCTGCAATGCTCAATCTTTAGTTGATGACAGATGCTTAC 4300
QY 1674 agcgacacagatagagaaagatggaaatgcgattgggagatgcatattttgcaggagaa 1733

```



Db	4299	AGGCACACAGATAGAGAAAGATGGAATGCGATTAGGATGCATATTTTCAGGACAGAA	4240	Medical Center Dr, Rockville, MD 20850, USA
QY	1734	cgcactgagctgtagaataatcctaatacctaatacgaatacattgattcagc	1793	Location/Qualifiers
Db	4239	CGCTAGTGAGCTGTAGAATAATGCTAAATCTAATAACATATAGCAACATAGATTACGC	4180	1. 12548
QY	1794	taagcagcaaatcgctactgactgaagacacacaaatagctgaagctcagaaaaagtcccccga	1853	/organism="Chlamydomophila pneumoniae AR39"
Db	4179	TAAAGCAGCAATCGCTACTGCTAAGACACAAATAGCTGAAGCTCAGAAAAAGTTCCCGCA	4120	/db_xref="AR39"
QY	1854	ctctcaaatcttcaagaagcgggacaacaaatggtatacacaaggctgagaagatcttaaaaa	1913	/note="synonym: Chlamydia pneumoniae AR39"
Db	4119	CTCTCCAAATCTTCAAGACAGCGAACAATAGTAAATACAGCTGAGAAAGATCTTAAAAA	4060	297. .965
QY	1914	tatcaaacctgcagatggtctcgtatgttccaaatccaggaactacagttggagctccaa	1973	/gene="CP0011"
Db	4059	TATCAACCTGCAGATGGTCTGATGTTCCAAATCCAGAACTACAGTTGGAGGCTCCAA	4000	297. .965
QY	1974	gaacaagaagtagtattggttagtattcggtttccatcgctgtagatgattcacaatcgga	2033	/gene="CP0011"
Db	3999	GCAACAGGAAGTAGTATTGGTAGTATTGCTGTTCCATGCTTGTAGATGATGCTGAAAA	3940	/note="similar to SP:P31218 GB:X71492 PID:296947 GB:U00096 PID:1736770; identified by sequence similarity; putative"
QY	2034	taagaccgttcattttatgctggtttgctgagatgattcacaatcgga	2093	/codon_start=1
Db	3939	TGAGACCGCTTCATTTTGTGCTGGGTTGCTCAGATGATTCATATACGGA	3880	/product="uridine kinase"
QY	2094	aaatctgattctcaagctgcccacagagctcgcagcacagctagagcagcgaagc	2153	/db_xref="GI:7188950"
Db	3879	AAATCTGATTCTCAGCTGCCACAGGAGCTCGCAGCACAGCTAGACGCGAAGC	3820	/translation="MLMMMLIIGTGGGAGKTTLTQNIKEIFGEDVSVICQDNYRK DRSHYTPERANLWDHPDAFDNLLISDIKRLKNEIWAQVDFVILGNSKTEIET IYPSKVLVEGLVFNQELRDLMDIRIFVDTDADERILRRMVRDVGQSDVDCIMS RYSLVKPMHEKFIETPKYADILVHGNYRQNVYTNILSOKIKHLENALBSEDTYYM VNSK"
QY	2154	cgtgagatgacagtgcgtcgcagcgtggtgagatgctcagaagctttagaagcggc	2213	complement(1047. .2021)
Db	3819	CGCTGAGATGACAGTGCCTGTCAGCGCTGGCAGATGCTCAGAAAAGCTTAGAAGCGCG	3760	/gene="CP0012"
QY	2214	tctagtaagctggcacaacagcgcatactcaatgcttggacagatcgtcttcgc	2273	complement(1047. .2021)
Db	3759	TCTAGTAAAGCTGGCCACACAGGCGGCTACTCAATGCTTTAGGACAGATCGCTTCGC	3700	/note="conserved hypothetical protein; identified by Glimmer2; putative"
QY	2274	tgtgtgtgagcgcaggag-tctctcccgctgcagcaagtctctatg	2319	/codon_start=1
Db	3699	TGCTGTGTGAGCGCAGGAGTTCTCCCGCTGCAGCAAGTTCTATAG	3653	/product="conserved hypothetical protein"
RESULT	3			/protein_id="AAF37908.1"
AE002165	12548 bp	DNA	BCT	/db_xref="GI:7188951"
LOCUS	Chlamydomophila pneumoniae AR39, section 2 of 94 of the complete			translation="MEKXYALAYYYTRVDNPHKEIALHKKFLEDLDVSCRIVISEQ GNGQFSGEPHEADYIMOWLERKPNFSAIKRIHIIKENIFRITVTKRKELALGCE VLSQAKHISPEQWEHEKQLQENRCLILDVNNYEWKIGFDNATLPDITOTREFPEYA EKLAQCDPRTTAVMYGTGICELYSVLLEKGFKEYQLDGVIAYGQVGTGKW LGKLFVDFDRLLAIPIDESPDVAPIAECCHQCTPSDVAANCANTCNALFLCCDCBIH QHGCCGECSCSPRVKFDSSRGKPFRAHLCEISENSEASACLL"
DEFINITION	genome.			/gene="CP0013"
ACCESSION	AE002165			2238. .2867
VERSION	AE002165.1			/gene="CP0013"
KEYWORDS	Chlamydomophila pneumoniae AR39.			/note="similar to SP:P81288; identified by sequence similarity; putative"
SOURCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.			/codon_start=1
ORGANISM	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			/product="ribosomal protein S4"
REFERENCE	1 (bases 1 to 12548)			/db_xref="GI:7188952"
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			translation="MARYCGPNRVARRFGANIFGRSRNPLLLKKPPQCHQMQRKK SDYGLQLEKOKLACYGMKEKOLVAFKEVHKGNVQAFMFLRPFRCRLDNMYRM GFAKTIFAAQQLVAGHILLVNGRVDRESFLRPGMOISLKEKSLQSVKDALESKD ESSLPSYISLDTGFKGELLVSPQDQIAQLPLFINISVCEFLSHRT"
TITLE	Nucleic Acids Res. 28 (6), 1397-1406 (2000)			complement(2986. .3867)
JOURNAL	20150255			/gene="CP0014"
MEDLINE	10684935			complement(2986. .3867)
PUBMED	2 (bases 1 to 12548)			/gene="CP0014"
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			/note="similar to SP:P12638 GB:M22591 PID:146954 PID:405898 GB:U00096; identified by sequence similarity; putative"
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			/codon_start=1
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			/product="endonuclease IV"
JOURNAL	20150255			/transl_table=1
MEDLINE	10684935			/protein_id="AAF37910.1"
PUBMED	2 (bases 1 to 12548)			/db_xref="GI:7188953"
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			translation="MKVLPPPSIPLLGAHTSTAGGLKNAIVEGRDIGASTVOIFTANO RQWRRLKKEVIEDFAALKETDLSYIMSHAGYLINFCADPPVILERSIGIYEIL DCITIGISFVNFPGCAALKSKEDMKNKIVSFSQAPLSPPLVLETTAGQGT LIGSNFEELGVLONKQIPIGVCVDTCIIFAGYDITSQGWEDVNEFDEYVGLS YLRAFLHNDSMFPLGANKDRHAPLGEVIGRESFKFLMTDERTKIPKYLETPGPGPEN WQKEIGELLKFSKNRDS"
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE	10684935			
PUBMED	2 (bases 1 to 12548)			
REFERENCE	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			
AUTHORS	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39			
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			
JOURNAL	20150255			
MEDLINE				



```
gene
complement(4202..4378)
/gene="CP0015"
CDS
complement(4202..4378)
/gene="CP0015"
/notes="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF37911.1"
/db_xref="GI:718954"
/translation="MSFLLKPSPPKDYLFDISPYTSBITIGGSYFKLNKASL
QSSTLRRSIIIS"
4362..6005
/gene="CP0016"
4362..6005
/gene="CP0016"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37912.1"
/db_xref="GI:718955"
/translation="MSRKDEVSILARSIFNLSCTFCRITGIFRETAMATYFADPI
VAAPWLGRTVFLKRLIGLLLEQAFIPHEFRRAQSLDRAAFRRFRSLKGSFI
IFTLLIEAVLVLIQVVEEGYDMILLTLLPCGIFLMMNVNGLHCENKFFGVG
LAPVYNIWIFFVIAHRSDPRRIIGLSVALVIGFFEWLIITVPGWVWKFLEAKSP
PQGHVSALLAPSLGLTSSIFQNLNLSICLIRVHEIGPLMYSLKIQVLPFH
LFGFVEVLLPATSRVOREHERGLKMLKLVLTMSVMIMTAGLLIALPGVRV
LYEHLFQSAVAIVRLVCGASIIIPMALAPLVSLVFLYAOROYAVPLFTGIGTALA
NVLSTLWGRWLKDVSGISATSIWVQLYFLWYSSKELPMYSLKILNESIRRSIK
NIGTLMACMITLGNILTOITYVIFLNPPLPLAWSSITAAQIAFLSESCIFLAF
FGFAXLREVDLINLASFEYWRGQGLLORHVQMDTON"
6108..7376
/gene="CP0017"
6108..7376
/gene="CP0017"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37913.1"
/db_xref="GI:718956"
/translation="MKQVYQWLASVLLALITSGYAEPLSEOKVSKHTVTTLDEVK
DYLKRGVETRRKQGVLRKIGADVRRWLYFREDIKNPSDKDKYNLPVNYRSEFYL
YIDYRAERNWLSKMNWTAAGGENTAAGVDINRAFLGRYKPNPRTDFMEIGRS
GLGDLFSEVQPSNFDGLHIYWTRELSKDYQYQVIVHGGPFVNMTHKHYAWVVEGI
LNLKPKQFVKCSVDWNTFVPSSTTEKATNAMKYKVCWQWLVGKHGSPWING
OKPLLYGAFILMNLAKATKTTLNGKENLAWF IGTTGLGLRKAGDWSATVRYEVEA
LSVPEDVSGIGRGLLKWFAQAIYAANDPKEANGFTNYKGFSAIYWGVTDSLSFR
AYGAVSKPANDKLGSDTFPRKFDLGIISAF"
7488..9443
/gene="CP0018"
7488..9443
/gene="CP0018"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37914.1"
/db_xref="GI:718957"
/translation="MVNPIGPGPIDETERTPPADLSAOGLEASAANKSAEAOIRIAGAE
APKPKESRVSERWLSILRSAYNALMSLADKLGIASSNSSSTSRANDVDSSTATPTP
PPPTFDYTKQATYDTFTSTSLADIQAALVLSQDAVINIKDPADEETIAAEW
ETKNADAVKVGQAQITELAYSDNQALDLSGLKTSFDLLQALLOSVADEENKAAELL
KEMQDNVPVPGTPAQAQSLVDQTDATQIEKDQNAIDFAFGAGNAGSAYENAKSN
NLSNIDSKAAATATAKTQIAEAKKFPDPSIQEAEQWVIAOEKDKLNKIPAGSDV
PNPCTVVGSKQGSIGSIRVSMILLDAENETASTLMSGFRQIMHMFNTENPDSQA
QBELAQAARNAKAGDSDAAALADAQKALEALGKAGOGGILLNALGQIAASAAVVA
GVPPAAASIGSSVKQLYKTSKSGSDYKIQISAGVDAYSINDAYGRANDATRDVI
```

```
Query Match      83.4%; Score 2122.2; DB 1; Length 12548;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2136; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

Qy 174 catgacaaaaaacattatgcttgggttgtagaggattctcaatgcttgctgaataa 233
Db 6722 CATGACAAAAAACATTATGCTTGGTGTAGAGGGATTCTCAATGTTTGCCTAAACA 6781

Qy 234 gtttttggtaaatgtagtgtgtcgactggagacacacattcgcttcttcagaaaaacccac 293
Db 6782 GTTTTGTGAAATGTAGTGTGTGCGACTGGAACACATTCGTTCTTCAGAAAAACCTCCAC 6841

Qy 294 tacgaaaaagcgtctacaaacgctatgaaatacaataactgtgtgttgagcagtgctcgt 353
Db 6842 TACGAAAAAGCCTGCTACAAACGCTATGAATAACAATACTGTGTGTGGCAGTGGCTCGT 6901

Qy 354 cggaaagcatagtcaggttccttggatcaatgagacagaaaaagcctctatatctttatgg 413
Db 6902 CGGAAAGCATAGTCAGGTTCCTTGGATCAATGGACAGAAAAAGCCTCTATATCTTTATGG 6961

Qy 414 agctttctaatgaaccccttagcaagcctacgaagactacgttaaatggaaagaaaa 473
Db 6962 AGCTTTCCTAATGAACCCCTTAGCAAGCCTACGAGAGCTACGTTAAATGGAAAGAAAA 7021

Qy 474 cctagcttggtttatttgaggaaacttttaggggagactcagaaagctgagagactggcgtgc 533
Db 7022 CCTAGCTGGTTTATTGGAGGAACCTTAGGGGGACTCAGAAAGCTGAGAGCTGCTCTGC 7081

Qy 534 cacagtcagttatgtagtgcgaagccttgcagttccagaaaaatagatgtttcaggaat 593
Db 7082 CACAGTACGTTATGAGTATGTCGAAGCCCTTGTCTGGTCCAGAAATAGAGATGTTTCAGGGAT 7141

Qy 594 tggccgtggttaattattaaagtgttgcgcccaagcaattgctgctaactatgatcc 653
Db 7142 TGCCCGTGGTAATTTATTAAAGTTTGGTTCGCCCAAGCAATGCTGCTACTATGATCC 7201

Qy 654 taaaaggcgtaatagttttacaaaataaaaggattttccgctctatatatgatgtatggat 713
Db 7202 TAAAGAGCGTAATGGTTTTCACAAATTAAAGAGATTTTCCGCTCTATATATGATATGGCAT 7261

Qy 714 cacagattctctatcattcagagcttatgggcttactccaaacagcaaacgataaact 773
Db 7262 CACAGATTCCTATCATTACAGAGCTTATGGGCTTACTCCAAACAGCAAGATAACT 7321

Qy 774 cggcagtgatttactttccgaaaagtgtgatagataatttcagcggttttaagtcac 833
```

```
gene
complement(4202..4378)
/gene="CP0015"
CDS
complement(4202..4378)
/gene="CP0015"
/notes="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF37911.1"
/db_xref="GI:718954"
/translation="MSFLLKPSPPKDYLFDISPYTSBITIGGSYFKLNKASL
QSSTLRRSIIIS"
4362..6005
/gene="CP0016"
4362..6005
/gene="CP0016"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37912.1"
/db_xref="GI:718955"
/translation="MSRKDEVSILARSIFNLSCTFCRITGIFRETAMATYFADPI
VAAPWLGRTVFLKRLIGLLLEQAFIPHEFRRAQSLDRAAFRRFRSLKGSFI
IFTLLIEAVLVLIQVVEEGYDMILLTLLPCGIFLMMNVNGLHCENKFFGVG
LAPVYNIWIFFVIAHRSDPRRIIGLSVALVIGFFEWLIITVPGWVWKFLEAKSP
PQGHVSALLAPSLGLTSSIFQNLNLSICLIRVHEIGPLMYSLKIQVLPFH
LFGFVEVLLPATSRVOREHERGLKMLKLVLTMSVMIMTAGLLIALPGVRV
LYEHLFQSAVAIVRLVCGASIIIPMALAPLVSLVFLYAOROYAVPLFTGIGTALA
NVLSTLWGRWLKDVSGISATSIWVQLYFLWYSSKELPMYSLKILNESIRRSIK
NIGTLMACMITLGNILTOITYVIFLNPPLPLAWSSITAAQIAFLSESCIFLAF
FGFAXLREVDLINLASFEYWRGQGLLORHVQMDTON"
6108..7376
/gene="CP0017"
6108..7376
/gene="CP0017"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37913.1"
/db_xref="GI:718956"
/translation="MKQVYQWLASVLLALITSGYAEPLSEOKVSKHTVTTLDEVK
DYLKRGVETRRKQGVLRKIGADVRRWLYFREDIKNPSDKDKYNLPVNYRSEFYL
YIDYRAERNWLSKMNWTAAGGENTAAGVDINRAFLGRYKPNPRTDFMEIGRS
GLGDLFSEVQPSNFDGLHIYWTRELSKDYQYQVIVHGGPFVNMTHKHYAWVVEGI
LNLKPKQFVKCSVDWNTFVPSSTTEKATNAMKYKVCWQWLVGKHGSPWING
OKPLLYGAFILMNLAKATKTTLNGKENLAWF IGTTGLGLRKAGDWSATVRYEVEA
LSVPEDVSGIGRGLLKWFAQAIYAANDPKEANGFTNYKGFSAIYWGVTDSLSFR
AYGAVSKPANDKLGSDTFPRKFDLGIISAF"
7488..9443
/gene="CP0018"
7488..9443
/gene="CP0018"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37914.1"
/db_xref="GI:718957"
/translation="MVNPIGPGPIDETERTPPADLSAOGLEASAANKSAEAOIRIAGAE
APKPKESRVSERWLSILRSAYNALMSLADKLGIASSNSSSTSRANDVDSSTATPTP
PPPTFDYTKQATYDTFTSTSLADIQAALVLSQDAVINIKDPADEETIAAEW
ETKNADAVKVGQAQITELAYSDNQALDLSGLKTSFDLLQALLOSVADEENKAAELL
KEMQDNVPVPGTPAQAQSLVDQTDATQIEKDQNAIDFAFGAGNAGSAYENAKSN
NLSNIDSKAAATATAKTQIAEAKKFPDPSIQEAEQWVIAOEKDKLNKIPAGSDV
PNPCTVVGSKQGSIGSIRVSMILLDAENETASTLMSGFRQIMHMFNTENPDSQA
QBELAQAARNAKAGDSDAAALADAQKALEALGKAGOGGILLNALGQIAASAAVVA
GVPPAAASIGSSVKQLYKTSKSGSDYKIQISAGVDAYSINDAYGRANDATRDVI
```



Db	7322	CGCGAGTGATTTTACTTTCGGAAGACTTGATGCTAGGTATAGGTATAATTTACAGCGTTTAAAGTCAA	7381
Qy	834	attttaaaaaatctttaaaaaacagcgtcgcattaaattatttagtgaagagcttttttttta	893
Db	7382	ATTTTAAATAAATCTTTAAAAACAGGCTCGCATTAATATTATTAGTGAGAGCTTTTTTTTTTA	7441
Qy	894	ttttttataaaactaaagattttttatt	953
Db	7442	TTTTTTTATTAATAAATCTAAAGATTTTTTATTATTATTTTTTTGAGTTTTTATGTTTAATCCTAT	7501
Qy	954	tggctccaggttccttatagacgaacagaacgcacacctccgcagatctttctgtccaag	1013
Db	7502	TGCTCAGGTCCTATAGACGAACAGACGACACACCTCCCGCAGATCTTCTGCTCAAG	7561
Qy	1014	atttgaggcgagtcgagcaaaataaagagtcggaagcctcaaaagatacagagtcggaagc	1073
Db	7562	ATTGGAGGCGAGTGCAGCAAAATAAGAGTGCAGAGCTCAAGAAATAGCAGGTGCGGAAGC	7621
Qy	1074	taagcctaagaatcttaagaccgattctgtagagcgatggagcactcttgctgtctgcagt	1133
Db	7622	TAAGCCTAAGAATCTTAAGACCGATTCTGTAGAGCGATGGAGCATCTTGGTCTTCGAGT	7681
Qy	1134	gaatgctctcatgagtcctggcagataaagctgggtattgcttcttagtaaacagctogtcttc	1193
Db	7682	GAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTTCTTAGTAACAGCTCGTCTTC	7741
Qy	1194	tactagcagatctgcagacgtggactcaacgcagcagccagccactacgctcctccacc	1253
Db	7742	TACTAGCAGATCTGCAGAGCTGGACTCAAGCAGCAGGACCGACCTACGGCTCTCCACC	7801
Qy	1254	cagctctgatattataaagactcaagcgcgaacagcttaecgatactatctttaccctcaac	1313
Db	7802	CAGGTTTGATGATTATAAGACTCAAGCGCAACAGCTTACGATACTATCTTTACCTCAAC	7861
Qy	1314	atcacctagctgcatacagctgcttggtagagcctcagagatcctgaactatgaataaaa	1373
Db	7862	ATCCTTAGCTTGACATACAGGCTGCTTGGTAGGCTCCAGGATGCTGTCTACTAATATAAA	7921
Qy	1374	ggatacagcggtctactgatgaggaacccgaatcgctgcggagtgggaactaagaatgc	1433
Db	7922	GGATACAGCGGCTACTGATGAGGAACCCGAATCGCTGCGAGTGGGAACCTAAGAATGC	7981
Qy	1434	cgatgcaattaaagtggcgcgcaaatcacgaattagcgaataatgcttcggataacca	1493
Db	7982	CGATGCAGTTTTAAGTTGGCGCCAAATTACAGAAATTAGCGAAATATGCTTCGGATAACCA	8041
Qy	1494	agcgattcttgactctttaggtaaactgactcctctgacacctctacagactgctctct	1553
Db	8042	AGCGATTCTTGACTCTTTAGGTAAACTGACTTCTCTTCGACCTCTTACAGGCTGCTCTCT	8101
Qy	1554	ccaatctgtagcaacaataacaagcgctgagcttcttaagagatgcaagataaacc	1613
Db	8102	CCAATCTGTAGCAACAAATTAACAAAGCAGCTGAGTCTTTAAAGAGATGCAAGATAACC	8161
Qy	1614	agtactccagggaaaaacccctgcgaatgctcaactcttagttgacagacagatgctac	1673
Db	8162	AGTAGTCCCGAGGAACCCCTGCAATTTGCTCAATCTTTTAGTTCATCAGACAGATGCTAC	8221
Qy	1674	agcgacacagatagagaagatggaagtgcgattggggatgcatttttcaggagacagaa	1733
Db	8222	AGCGACACAGATAGAGAAAGATGGAAATTCGCAATTAGGGATGCATATTTTTCGAGGACAGAA	8281
Qy	1734	cgctagtgaagctgtagaaaatgctaaatcctaataacagtaataagcaacatagattcagc	1793
Db	8282	CGCTAGTGAGCTGTAGAAAAATGCTTAATATCTAATAACAGTATAGCAACATAGATTACGC	8341
Qy	1794	taaagcagcaaatcgctactgtctaagacacaaatagctgaagctcagaaaaagttccccga	1853
Db	8342	TAAAGCAGCAATCGCTACTGCTTAAGACACAAATAGCTGAACCTCAGAAAAAAGTTCCCGA	8401
Qy	1854	ctctccaattcttcaagaagcggaaacaaatggtaataacagcgctgagaaagatcttaaaa	1913
Db	8402	CTCTCCAAATCTTCAAGAACGGAACAAATGGTAAATACAGCTCAGAAAGATCTTAAAAA	8461

Qy	1914	tatacaacctgcagatggttctgtatgttccaaatccaggaactacagtgtgagggtccaa	1973
Db	8462	TATCAAACTCTGCAGATGGTTCTGATGTTCCAAATCCAGGAACACAGTTGGAGGCTCCAA	8521
Qy	1974	gcaacaaggagtagtattgtagtatttcgtgtttccatgctgttagatgatgtcgaaga	2033
Db	8522	GCAACAAGGAAGTAGATTGGTAGTATTCTGTTTCCATGCTGTAGATGATGCTTGAAAA	8581
Qy	2034	tgagaccgcttccattttgatgtctctgggtttcgtcagatgattccatgtttcaatcagga	2093
Db	8582	TGAGACCGCTTCCATTGTTGATGCTCGGTTTCGTTCAGATGATTCACATGTTCAATACGGA	8641
Qy	2094	aaatcctgatattcgaagctgcccaacagagagctcgcagcaagaagtagagcagcgaagc	2153
Db	8642	AAATCTCTGATTCTCAAGCTGCCCAAGAGAGCTCCGACACAAAGCTAGACGACGCAAAAGC	8701
Qy	2154	cqctggagatgacagtgctgctgcagcgcctgcagatgctcgaagaagctttagaagcgc	2213
Db	8702	CGCTGGAGATGACAGTGTCTGCTGCACGCGCTGCAGATGCTCAGAAAGCTTTAGAAAGCGCG	8761
Qy	2214	tctaggtaagctgggcaacaacagggcactactcaatgctttgggacagatcgcttctgc	2273
Db	8762	TCTAGGTTAAAGCTGGGCAACAACAGGGCATACTCAATGCTTTAGGACAGATCGCTTCGCG	8821
Qy	2274	tctgttgtgagcagcaggag-tctctcccgctgcagcaagttctatgg	2319
Db	8822	TGCTGTGTGAGCGCAGGAGTCTCTCCGCTGCAGCAAGTTCTATAG	8868
RESULT	4		
LOCUS	AP002547/c		
DEFINITION	AP002547 300550 bp DNA BCT 08-DEC-2000		
ACCESSION	AP002547 AB033786 AB033787 AB033816 AB033817 BA000008		
VERSION	AP002547.2 GI:10176693		
KEYWORDS			
SOURCE	Chlamydothila pneumoniae J138 (strain:J138) DNA.		
ORGANISM	Chlamydothila pneumoniae J138		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothila.		
AUTHORS	Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.		
TITLE	Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CW029 from USA		
JOURNAL	Nucleic Acids Res. 28 (12), 2311-2314 (2000)		
MEDLINE	20330349		
REFERENCE	2 (bases 1 to 300550)		
AUTHORS	Shirai, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.		
COMMENT	Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail: mshirai@epo.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227, Fax: 81-836-22-2415)		
FEATURES	On Sep 15, 2000 this sequence version replaced gi:6172298		
source	gi:61722400 gi:6172396 gi:6172398 gi:8978889.		
source	AB033786-AB033787, AB033816-AB033817; Submitted (25-Oct-1999).		
source	Location/Qualifiers		
source	1. .300550		
source	/organism="Chlamydothila pneumoniae J138"		
source	/strain="J138"		
source	/db_xref="taxon:138677"		
source	/note="synonym:Chlamydia pneumoniae (strain J138)"		
source	complement(21..794)		
source	/genes="dapF"		
source	complement(21..794)		
source	/genes="dapF"		
source	/codon_start=1		
source	/transl_table=11		
source	/product="diaminopimelate epimerase"		
source	/protein_id="BAA98725.1"		



```

/db_xref="GI:8978890"
/translation="MAFISPTSTKSIYKAGNRLLGLTLEVEDVDFLCOETVRD
GFLYKPSDAQQLIIFNSDGRPTMCGNLCAIAHLASQKGSIDISVSTDSGLYS
GYFYSWRLVDMTADWRASVHRLESRPDLPEKVCVCHTGPVHAVVILPEISTDL
SILGFLRYHTSPDPCNVNFVQILGHCOLRVYTVRGEVGETAACGTGALASALVV
SNSWGKESIOIHWGGLMELTVSQNRGVYLGQSVTRDL"
complement(763..1338)
/gene="clpP_1"
/complement(763..1338)
/gene="clpP_1"
/codon_start=1
/transl_table=11
/product="CLP protease"
/protein_id="BAA98726.1"
/db_xref="GI:8978891"
/translation="MADGEVHKLRDIEKELLEARRVFFSEPVTEKSASDAIKKLWYL
ELKDPGPVIVINSPPGSDVAGFAVMDQIKMLTSPVTVTVTGLAAGMGSVLSICAAP
GRRATPHSRIMIHPSIGPITGOATDLDIHAREILTKARIIDVVVEATNQPRDII
EKAIDRWMTANAKDFGLDGLIFSENDL"
complement(1358..2917)
/gene="glyA"
/complement(1358..2917)
/gene="glyA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/translation="MLKVFKEFKFAIVEIFTKVAVVSLHLKLENASKKGOSLAS
TAYLAADHLNAPFSGIERIIDELKQSRSHUKMIASENYSLSVQAMGNLITDKYC
EGSPKFRYSCCENVDATMECVETAKELFACACVQPHSGADANLAVMAILTHKV
OGPAVKLGKTVNELTETETLLKAEBSFCCVCLGSLNSGGHITGNVRLNVMSKLM
RCFVDPNPDECPDYAEISLRAKEYPKVLIAGYSVSRRLNAVLKQIAEDCGSVL
WYMAHFAGLVAGGVFVDEENPIYADIVTTHKTLLRGPGLVLAETREYESTLNG
CPLMGGPPIVIAKTAVALKEALSVDFFKKAHQVNNARLAERFISHLRLITGGT
GNMIVDILGSLGTSKTAEDILSSVGIANNRSLPSDAIGKWDTSGLRGITGPALTIL
GMIIDEEVEADIIYKVLNRILRSCHVEGSKKNGKELPEAIAQEARDRVRNLLRFP
LYPETIDLEALV"
3107..3778
/gene="CPj0522"
3107..3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="Ct433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/translation="MTLXILGNQKTARKYQAHYLPILTLFPYAKSTPONKRALQFLPQ
ATHVILTSPSSHFLSRMTSLLSKATLTKTYLIGESTKERLLSLFGQVKYVYATQ
ETAEIGIFLLQALPSSARILYPHSSLARPVIREFLYNRETFSPHYTVKPRKLKNI
LSYKKKIIFTSPTVRAPAKIFPRPEPTYWCQGRMTLQEPKFTSSQKQVSLLETGK
SRTSP"
3843..4175
/gene="CPj0523"
3843..4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/translation="MASSATPGFDGTASLPFPATPRPNKFLAFVITAIATLVWIAL
IATTAIGICHIPLCSFLTAIFLYISRYTCSYARNVYIALDVVPDHSKLQDMRS
HSPIFSDR"
4233..5302
/gene="CPj0524"
4233..5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"

```

```

/translation="MSGPSRTESSQSVLSYVPRDKIEAPKKQFTIAKISTLAILASL
ALGAVAGISLTVILGNPVFALLITFALFVSFVTVFVYHMTSKVSNQKVKLEQNFK
PLGKAWCNCVQSYNEMQFYNNHLNPKFAVIAOTDASQPOFTFLTLGRVTEKQNST
GIIFNPVGTNLDINTATNLSTILSTLKDSVWOTCQREGGPAKGDPPSPTEVRV
VKLPNDQDTFNLNLSSAEKSSILPTFLGHVCGSPKSELPNQOETIYKQALLAYENCL
KAAIESHAAIVALPLFTSYEVPPEEILPKEGTFYWDNQTOAFCK3ALLDAIONTALR
YFQSRLLVILQDPFNTIESQRSEE"
5642..6406
/gene="CPj0525"
5642..6406
/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="Ct398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/translation="MHDALSILAIQELDIKIMLRMRVKKHKEQELAKVQSLKSDIRR
KVQEKLEMLNKTQIRDGENRIQETSEIQINKLENOAAVKMKDEFNALTQEMTANK
ERRSLEHQLSDLMKQAGGEDLIVKSLASTENSSSVIEREIESIKKINBEKCAL
LEORTLKHATNPELLSIYERLLNNKDRVVPVPIENRVCSCGCHIVLTPQHENLVRRKD
RLIFCEHCSRILYQESQVNAQENSTAKRRRRRAV"
complement(6827..7816)
/gene="kpsF"
/complement(6827..7816)
/gene="kpsF"
/codon_start=1
/transl_table=11
/product="GutO/KpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/translation="MPSMISTDVCODILIGKEAVDFFQAFQPKENQLAKEKILGH
SGWFTSGVGKSCVARKLVATLQISLERALFFSPVDLLHGDGLVSPGDIVCLFSKS
GETQELIDTVPLHLSRAILVAITSPYLNLAASDLVILPVSADLPFNLIPTNST
TCOMIGEDFLAMLLFHSRGLVSTYGNKHPSPGOVGMKANGKVDKPEFTEVFCFLG
DKVSELEVSFAVCGCVCIVDPQFLMGLFTDGLRRLSLASVGGVGLSLSLEKVMTA
NPRCIEDSDIALQLMSESSSPVAVLPVLNDENRHTVLTGLLHMHTLAKAGLL"
complement(7843..9024)
/gene="sucB_2"
/complement(7843..9024)
/gene="sucB_2"
/codon_start=1
/transl_table=11
/product="dihydrodiploamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/translation="MIFEFPRKIGETSSGGSIVRMLKNLGHVARDPELIEVSTDKI
ATELPSKAGRLVRCFVNEGDEVASGDLVGLIELEETSEADDETSCLPTSCETKSEA
GSSSSVWFSPAVLSLAOREGIGLNDLQIAGTKGGRVTRDLEAVISESQOVSPIE
IFQGEVNRIPMSPLRRAIASLSKSDSEVPHASLVVDVDTDLMLNLSIGSRQRFLOTH
GVKLITTSFIVQCLAQRLRQFPLNGSLDGTITVMKSVNGVAVNLNKEGVVVPVTH
NCODRGLVSTAKALADLSSRARLNKLDPSEVQDGSVTVTNFGMTGALIGMPIIRYPEV
AILGIGTIQRVVVVRDDDSLAIKRMVYVTLTFDHRVLDGIYGEFUTSLKNRLESVTM
G"
complement(9038..10282)
/gene="gltr"
complement(9038..10282)
/gene="gltr"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/translation="MKLWMKIFGLGVGTGLVLEDKAIFPKFIGDIFNLNLSMVVY
PLVFCSSVGLIASISDMKKLGRIGISIAQVFPNPNVRSFAEGNLIQIILFAIFGLALRL
QAQSDMSAVTVIDSNTAAVFLSIIAQVFPNPNVRSFAEGNLIQIILFAIFGLALRL
SGERGPRVERIDGSEITMLRMVNMIMSFPYGVGASMAWISGNHGLVGLWQLGKFI
AYVLACLPHATLVEGLVFCGKMSFKSLSSMDAISCAVSTASSATLPVTMRCVS
KNLGVSAFVSGVFLPLGATVNMNGTAIFQGMAAVFIQAQVNCPLSLSLILLVVTATF
SAGSAGVPGGGMITLGSVASVGLPIQGIALLAGIDRLDIVCTPWNILGDVAVVATY
VASEGELSYESIKQSEVET"
complement(10282..11379)
/gene="ycaH"
complement(10282..11379)

```



```

/gene="ycat"
/codon_start=1
/transl_table=11
/product="ATPase"
/protein_id="BAA98735.1"
/db_xref="gi:8978900"
/translation="MKRRPSTLFLFYRRVITAIISLEGILGMWGLSLLSKVFAFLVA
CWNRFSTPYRARSSTVISGNIIVGGAGKTPTLWLAELRLRGYSGVLSRGYKSO
SSRQKLTVDKSVHSASYVGDEPLLAELPEGSVWVKDRRIISAARAAKFGILL
DDGLQYRLKHDVLAIVNGDPLGGRAFFPKRLDPLRLKTVDAITLVNGGKEAG
TVARVSNAPOLPVKPTIASVYVTHNGERIPKEALRELRVGFGGLFPQGLPNLRE
EGHILSKYLDPDHAATKKELTNQQQGGGLCTEKSVKPLRUSGEVSLIP
IAKEMRLSYNQDDTLNLLNIEQIHKNRGN"

```

Query Match		83.4%	Score 2122.2;	DB 1;	Length 300550;
Best Local Similarity		99.6%	Pred. No. 0;		
Matches	2138;	Conservative	0;	Mismatches	8;
				Indels	1;
				Gaps	1;
QY 174	catgacaaaaaacattatgcttgggtgttagaaggattctcaatcgcttgcctaaaca	233			
Db 218364	CATGACAAAAAACATTATGCTGGGTGTAGAGGATTCCTCAATCGCTTGCCTAAACA	218305			
QY 234	gttttttgaaatagttgttcgactggaacattcgcttcctcagaacacctcaac	293			
Db 218304	GTTTTTTGAAATGTAGTGTGTGCTGACTGGAACACATTCGTTCTCAGAAAACCTCCAC	218245			
QY 294	tacagaaaaagctctacaaacgctatgaatacacaatactgtgttgagcagtgctcgt	353			
Db 218244	TACAGAAAAGCTGCTACAAACGCTATGAAATACAAATACTGTGTTGGCAGTGGCTCGT	218185			
QY 354	cggaaagcatagtcaggttcttggatcaaatggacagaaaaagcctctatatctttatgg	413			
Db 218184	CGGAAAGCATAGTCAGGTTCCTTGGTCAATGGACAGAAAAAGCCTCTATATCTTTATGG	218125			
QY 414	agctttcttaagaacctttagcaaggctcgaagactacgtttaaatggaaagaaaa	473			
Db 218124	AGCTTCTTAATGAACCTTTAGCAAGGCTACGAAGACTACGTTAAATGGAAAAAGAAA	218065			
QY 474	cctagcttggtttattggaggaaactttaggggactcagaagaagctggagactggtctgc	533			
Db 218064	CCTAGCTGGTTTATTGGAGGAACTTTAGGGGACTCAGAAAAGCTGGAGACTGGTCTGC	218005			
QY 534	cacagtcagttatagtagtgcgaagccttgcagttccagaaaaatagattgttcaggat	593			
Db 218004	CACAGTACGTTATGATAGTATGTCGAAGCCTTGTGCGTTCCAGAAATAGATGTTTCAGGGAT	217945			
QY 594	tggccgtggttaattattaaagtgttgcgcgaagcaattgctgctaactatgatcc	653			
Db 217944	TGGCCGTGTTAATTTATTAAGGTTTTGGTTCGCCCAAGCAATTCGCTTAATATGATGCC	217885			
QY 654	taaaagagcctaagttttacaaaattataaaggattttccgctctatatatgattggcat	713			
Db 217884	TAAAGAGGCTAATGGTTTTACAAATTAATAAGGATTTTCCGCTCTATATATGATGGCAT	217825			
QY 714	cacagattctctatcattcagagcttatgggcttactccaaacacagaaacgataaac	773			
Db 217824	CACAGATTCTCTATCATTACAGAGCTTATGGGCTTACTCCAACACAGCAACGATAAAT	217765			
QY 774	cggcagtgattttactttccgaaagtgttagctaggtataatttcagcggttttaagtcac	833			
Db 217764	CGGCAGTGATTTTACTTTCCGAAAGTTTGATCTAGTATTAATTTTCAGCGTTTTTAAGTCNA	217705			
QY 834	attttaaaaaattctttaaaaaacagcgctgcattattattagtgagagctttttttta	893			
Db 217704	ATTTTAATAAAAACTTTAAAAACAGCTCGCATTAATTAATTAAGTACGCTTTTAAATGTA	217645			
QY 894	tttttttaataaaaaaactaaagtattttatttttttttttttttttttttttttttt	953			
Db 217644	TTTTTTTAATAAAAACTAAAGATTTTATTAATTTTGGTGTATTTATGGTTAATCCCTAT	217585			
QY 954	tggctcaggtccttatagacgaacagacgcacacctcccgagatctttctgctcaagg	1013			
Db 217584	TGGTCCAGGCTCTATAGAGGAACAGACGCACCTCCCGCAGATCTTCTGCTCAAGG	217525			

QY 1014	attgagcgagtcgagcaataaagagtcggaagactcaaaagatagcaggtgcggaagc	1073
Db 217524	ATTGGAGCGGAGTGCAGCAATAAAGATGCGGAAGCTCAAGAATAGCAGGTGCGGAAGC	217465
QY 1074	taagcttaaaatcaagaccgattctgtagagcgatggagcatcttgcgttctgcagt	1133
Db 217464	TAAGCCTAAAGAAATTAAGACCGATTCGTACAGCGATGGAGCATCTTGCCTGCTGCACT	217405
QY 1134	gaatctctcatgaactcggcagataaagctgggtatttgcctctagtaaacagctcgtcttc	1193
Db 217404	GAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGTAACAGCTGCTCTTC	217345
QY 1194	lactagcagatctgcagactgagctcaacgacagcgaccgacactacgcctctccacc	1253
Db 217344	TACTAGCAGATCTGCAGACGTGGACTCAACGACAGCGACCGACCTACGCCTCTCCACC	217285
QY 1254	cagctctgatgattataagactcaagcgcaaacagcttacgatactatttaccctaac	1313
Db 217284	CACGTTTGTATGATTATAAGACTCAAGCGCAACACAGCTTACGATACTATCTTTACCTCAAC	217225
QY 1314	atcactagctgacatacaggtcgttttggtgagcctccaggtgctgctcaactaataaa	1373
Db 217224	ATCAGTAGCTGACATACAGGCTGCTTGTGGAGCCTCCAGGATGCTGTCACTAAATATAA	217165
QY 1374	ggatacagcgctactgtaggagaaacgcgaatcgctcgagtgaggaaactcaagatgc	1433
Db 217164	GGATACAGCGGCTACTGATGAGGAACCGCAATCCGCTCGGAGTGGGAAACTAAGAAATGC	217105
QY 1434	cgatgcaattaaagtggcgcaaatcacagaattagcgaataatgctctcggaataacca	1493
Db 217104	CGATGAGTTAAAGTTGGCGCAAAATACAGAAATAGCGAAATATGCTTCGATAACCA	217045
QY 1494	agcagattctgactcttagttaaactgacttccttcgacctcttacagactgctctct	1553
Db 217044	AGCGATTCTTGACTCTTTAGGTAAACTGACTTCTTCGACCTCTTACAGGCTGCTCTTCT	216985
QY 1554	ccaatctgtagcaacaatacaaaagcagctgagctctcttaagagatgcaagataaac	1613
Db 216984	CCAACTGTAGCAACAATAACAAAGCAGCTGAGCTTCTTAAAGAGATGCAAGATAACCC	216925
QY 1614	agtagtcccagggaacacgcctgcaattgctcaattcttagttgtagcagacagatgctac	1673
Db 216924	AGTAGTCCCAGGGAACGCCCTGCAATTTGCTCAATCTTTAGTTGATCAGACAGATGCTAC	216865
QY 1674	agcgacacagatagagaaagatggaaatgcgattgggagtcataattttgcagggacaaa	1733
Db 216864	AGCGACACAGATAGAGAAAGATGGAAATGCGATTTAGGATGCTATTTTTCGAGGACAGAA	216805
QY 1734	cgctagtggagctgtagaaaatgctaaaatcctaataacagataaagcaacatagattcagc	1793
Db 216804	CGCTAGTGGAGCTGTAGAAAATGCTTAAATCTTAATAACAGTATAAGCAACATAGATTACG	216745
QY 1794	taaaagcagcaatcgctactgctaaagacacaaatagctgaagctcagaaaaagttcccgca	1853
Db 216744	TAAAGCAGCAATTCGCTACTGCTTAAGACACAAATAGCTGAAGCTCAGAAAAAGTTCCCGCA	216685
QY 1854	ctctccaattctcgaagcggaaacaaatgtaatacagcgtgagaagatcttaaaaa	1913
Db 216684	CTCTCCAATTTCTTAAGAAAGCGGAACAATAATGGTAAATACAGGCTGAGAAAGATCTTAAANA	216625
QY 1914	tatcaaacctgcagatggttcttgatgttccaaatccagaacactacagttggaggtcccaa	1973
Db 216624	TATCAAACTGCAGATGGTTCTGATGTTCCAAATCCAGGAACCTACAGTTGGAGGCTCCAA	216565
QY 1974	gcaacagaagaagtagtattgtagtattcggtttccatgctggttagatgagtgcaaaa	2033
Db 216564	GCACAAGGAAGTAGTATTGGTAGTATTTCGCTGTTTCCATGCTGTTAGATGATGCTGAAAA	216505
QY 2034	tgagaccgtctccattttgattgctggttttcgtcagatgattcacatggttcaatcacga	2093
Db 216504	TGAGACCGCTTCCATTTTGATGCTCTGGGTTTTCGTCGATGATTACATGTTCAATACGGA	216445



Qy 2094 aaatcctgattctcaagctgcccaacaggagctcgccagcaagctagagcagcgaaagc 2153  
 Db 216444 AAATCTGATTCTCAAGCTGCCAACAGGAGCTCGCAGCACAAGCTAGAGCAGCGAAGC 216385  
 Qy 2154 cactgagatgacagctgctcagcgctgagcgctgagctgagcagcagcttagaagcgcc 2213  
 Db 216384 CCGTGGAGATGACAGTGTCTGCTGAGCGCTGCGCAGATGCTCAGAAAGCTTTAGAAGCGGC 216325  
 Qy 2214 tctagtgtaagctgggcaacacaggcgctactcaatgcttctggagcagatcgcttctgc 2273  
 Db 216324 TCTAGGTAAGCTGGGCAACACAGGCGCATCTCAATGCTTTAGACAGATCGCTTCTGC 216265  
 Qy 2274 tctgtgtgagcagcagag-tctcccgctgagcagcttctatgg 2319  
 Db 216264 TGCTGTGTGAGCGCAGGAGTCTCTCCGCTGCAGCAAGTCTCTATAG 216218

RESULT 5  
 CH776KDA CH776KDA 2148 bp DNA BCT 17-MAR-1994  
 LOCUS Chlamydia pneumoniae 76 kDa protein gene, complete cds.  
 DEFINITION L23921  
 ACCESSION L23921  
 VERSION L23921.1 GI:435961  
 KEYWORDS 76 kDa protein.  
 SOURCE Chlamydia pneumoniae DNA.  
 ORGANISM Chlamydia pneumoniae  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 1 (bases 1 to 2148)  
 Perez-Melgosa, M., Kuo, C.-C. and Campbell, L.  
 Isolation and characterization of a gene encoding a Chlamydia  
 pneumoniae 76-kilodalton protein containing a species-specific  
 epitope  
 Infect. Immun. 62, 880-886 (1994)  
 JOURNAL  
 MEDLINE  
 FEATURES  
 source

Location/Qualifiers  
 1..2148  
 /organism="Chlamydia pneumoniae"  
 /db\_xref="taxon:83558"  
 1..2148  
 /codon\_start=1  
 /transl\_table=11  
 /product="76 kDa protein"  
 /protein\_id="AAA23117.1"  
 /db\_xref="GI:435961"  
 translation="MTKKHYAVVVEGILNRLPKOFFVKCSVVDWNTFVPSSTTERA  
 ATNAMYKYCVWOLVGHKSQVPHINGKPELYLGFALMPLAKATKTLINGENLA  
 WFIGTGLLRAGDWSATVREIVREIALSVFEIDVSGIRGNLLKFWFAIAANYDP  
 KEAGFTNYKGLSALYMGITDSLSFRAYGASKPANDKLGSDFTFRFDLGLISAG  
 KSNFNKIENKRLIISSEFFIYFKTKRELLFEFLVNPVIGPIDEPTPPAD  
 LSAOGLSAANKSAEQAIRIAGAEAKPKESKTSVERWSILRSVNALMSLADKLGA  
 SSNSSSSTSRADVDSTTATPTPPPTDDYKTAQOAYDTIFTSLADIQAALVS  
 LQDVTNIDKTAADDEETAIAEWETKNADAVKVAQITELAKYASDNOAILDLSGLK  
 TSFDLLQALQLGSVANNKNAELKEMQDNFVVGKTPFAIAQSLVDQTDATQIEXD  
 GNAIRDAYAGONAGAVENAKSNISIDSAKAAIATQIAEAKKFPDPSPILO  
 EAEQWVQMIKDKNIKPADGSDVNPQGTGSGKQSGSSIGSRVSMLLDADENETA  
 STLMGSQFQMIHMTENPDQSAQOELAAQARAARAKAAGDSDSAAALADAKALEAL  
 GRAGQOGLINALGOIASAAVVSAGVLPLOQVL"  
 BASE COUNT 659 a 435 c 480 g 574 t

Query Match 82.9%; Score 2109.2; DB 1; Length 2148;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 2137; Conservative 0; Mismatches 8; Indels 3; Gaps 2;  
 Qy 175 atgacaaaaaacattatgctgggttagagggaggttcaatgcttgcctaaacag 234  
 Db 1 ATGACAAAAAACATTATGCTGGGTGTAGAGGGATTCTCAATGCTTTGCTTAAACAG 60  
 Qy 235 tttttgtgaatgtagtctgagcagcagcagcagcagcagcagcagcagcagcagcagc 294  
 Db 61 TTTTGTGAAATGTAGTGTGTGCGACTGGAACACACATTGCTTCTTCAGAAACCTCACT 120

Qy 295 acagaaaaagctgctacaaacgctatgaaatcacaaactgctgtgttggcagtggtcgtcc 354  
 Db 121 ACAGAAAAAGCTGCTCAAAACGCTATCAAAATCAAAATCTGTGTGGCAGTGGCTCGTC 180  
 Qy 355 ggaagcatagtcaggtctccttggatcgaatgacagaaaaagcctctatatctttatgaa 414  
 Db 181 GGAAGCATAGTCAGGTCTCCTTGATCAATGGACAGAAAAAGCCTCTATATCTTTATGGA 240  
 Qy 415 gctttttaaagaaccccttagcaaaaggctacgaagaactacgttaaatgaaaaaagaaac 474  
 Db 241 GCTTTCTTAATGAACCCCTTTAGCAAAAGGCTACGAAGACTACGTTAATGGAAGANAAC 300  
 Qy 475 ctgactggtttattagggaaacttttaggggagctcagaaaaagctggagactggtctgcc 534  
 Db 301 CTAGCTTGGTTTATTGAGGAACCTTTAGGGGACTCAGAAAAGCTGGAGACTGGCTGCC 360  
 Qy 535 acagtcagttatgagtagtcgaagccttgcagttccagttccagaaaaatagattgttcaggatt 594  
 Db 361 ACAGTAGCTTATGAGTATGTCGAAGCCTTGTGCGTTCCAGAAATAGATGTTTCAGGSAAT 420  
 Qy 595 ggcgtggttaattattaaagtttgggttcgcccagcaattgctgtaactatgatcct 654  
 Db 421 GCGCGTGTAAATTAATAAGTTTGGTTCGCCCAAGCAATGCTGCTAACTATGATCCT 480  
 Qy 655 aaagagcctaagtagttttacaaattataaaggattttccgctctatatatgtatggcatt 714  
 Db 481 AAAGAGGCTAATGTTTACAAATTAATAAGGATTTTCCGCTCTATATATGTATGGCATC 540  
 Qy 715 acagattctctataatcagagcttgggttcggttactccaaacagcaaacagataaactc 774  
 Db 541 ACAGATTCTATCATTCAGAGCTTATGGGCGTTTACTCCAAACAGCAACAGATAAACTC 600  
 Qy 775 ggcagtgattttactccgaaagttagtctaggtataatttcagc--gttttaagtca 832  
 Db 601 GGCAGTGAATTTACTTCCGAAAGTTTGATCTAGGTATATTTTACAGCAGGTTTAAAGTCA 660  
 Qy 833 aattttaaaaaactttaaaacagcgtcgacttaattattagtgagagcctttttttt 892  
 Db 661 AATTTTAAATAAATCTTTAAACACAGGCTCGCATTAATTTATAGTAGAGCTTTTTTTTT 720  
 Qy 893 attttttaaaaaaagagattttatttttttgggttttta-tggttaactcct 951  
 Db 721 AATTTTATAATAAACTTAAAGATTTTATTTATTTTGTGATTTTATTTTGGTTAACTCT 780  
 Qy 952 attggtcaggtcctatagacgaacagacgcacacccctccgcagatctttctgctcaa 1011  
 Db 781 ATTGCTCAGGCTCTATAGACGAAACAGACGACACCTCCCGCAGATCTTCTGCTCAA 840  
 Qy 1012 ggaattggaggcgagtcagcaaaaataagagtcggaagctcaaaagataagcaggtcgcaa 1071  
 Db 841 GGATTGGAGGCGAGTTCGACGAAATTAAGAGTCCGGAAGCTCAAGAAATAGCAGGTGCGGAA 900  
 Qy 1072 gctaaagcctaagaactcaagacgattctgagcgatgagcattctgcttctgca 1131  
 Db 901 GCTAAGCCTTAAAGAACTTAACACGATTTCTGTAGAGCGATGAGCACTTTCGTTCTGCA 960  
 Qy 1132 gtgaatgctctcatgagtcgtggcgagataagctgggtattgcttcttagtaacagctcgtct 1191  
 Db 961 GTGAATGCTCTCATGAGTCTGCAGATAAGCTGGGTATTGCTTCTAGTAACAGCTGCTCT 1020  
 Qy 1192 tctactagcagatctgagcagcgtgagctcaagcagcagcagcagcagcagcagcagcagcagc 1251  
 Db 1021 TCTACTAGCAGATCTGCAGAGCTGAGCTCAACGACGAGCGACCGACCTTACGCCCTCTCCA 1080  
 Qy 1252 cccagctctgattataaagactcaagcgaacagcttacgatactatctttacctca 1311  
 Db 1081 CCCAGTTTGATGATTATAGACTCAACGGCAACAGCTTACGATACATCTTTACCTCA 1140  
 Qy 1312 acatcactagctgacatacagcgtgcttggtagcctccagagatgctgctcaataata 1371  
 Db 1141 ACATCAGTACGTACATACAGGCTGCTTTGGTGGAGCTCCAGGATGCTGTCTACATAATA 1200  
 Qy 1372 aaggatacagcggtcactgtagaggaacaccgcaatcgtcgaggagtggaacaaagaat 1431







```
QY 1380 agcgctactatagaaacccgaatcgctgcgagtgaggaaactaagaatgccgatgc 1439
DB 541 AGCGCTACTGATGAGAAACCGAATCGCTGCCGAGTGGAAACTAGAAATGCCGATGC 600
QY 1440 aattaagtctggcgcgaataatcagaattacgaattagcgaataatgtcttcgataaccacgat 1499
DB 601 AGTTAAAGTGGCGGCAATACAGAAATAGCGAAATATGCTTCGGATAACCAAGCGAT 660
QY 1500 tcttgactcttagtgaaactgaactctcttcgacctcttacagactgctcttccaatc 1559
DB 661 TCTTGACTCTTTAGTAAACTGACTTCCTCGACTCTTACAGCTGCTTCCTTCCTCAATC 720
QY 1560 tctacaaactaacaagcagctgagctcttctaaagagatgaagataaccacagtagt 1619
DB 721 TGTAGCAACAATAACAAGACGCTGAGCTTCTTTAAAGAGATGAAGATAACCCAGTAGT 780
QY 1620 cccagggaaacgcctgcaatctctttagttgattgacagacagatgctacagcgac 1679
DB 781 CCCAGGAAACGCCTGCAATTTGCTCAATCTTTAGTTGATCAGACAGATGCTACAGCGAC 840
QY 1680 acagatagagaagatggaatgcgattggtgggagatcatatttgcagacagaagcgttag 1739
DB 841 ACAGATAGAGAAGATGGAATGCGATTAGGGATGCATATTTTGAGGACAGAACGCTAG 900
QY 1740 tggagctgtagaatgctcaatctcaataacagatataagcaacatagattcagctaaagc 1799
DB 901 TGGAGCTGTAGAATGCTTAATCTAATAACAGTATAGCAACATAGATTTCAGCTAAAGC 960
QY 1800 agcaatcgctactgtctgaagacacaaatagctgaagctcagaataagttcccgactctcc 1859
DB 961 AGCAATCGCTACTGCTTAAGACACAAATAGCTGAAGCTCAGAAAAAGTTCCCGACTCTCC 1020
QY 1860 aattctcaagaagcgaacaaatggttaatacagctgagaaagatcttaaaatatcaa 1919
DB 1021 AATTCTCAAGAGCGGAACAATGTAATACAGCTGAGAGCTGAGAAGATCTTAAATAATCAA 1080
QY 1920 acctgcagatggtctctgattgtccaaatccaggaactacagttgaggtcccaagcaaca 1979
DB 1081 ACCTGCAGATGGTCTGATGTTCCAAATCCAGGAACACTACAGTTGGAGGCTCCCAAGCAACA 1140
QY 1980 aggaatgattggttagtattgtcttcctatgctgttagatgattgctgaaatgagac 2039
DB 1141 AGGAAGTAGTATTGGTAGTATTGCTGTTCCATGCTTTAGATGATGCTGAAATAAGAC 1200
QY 2040 cgctccatttgatgctcgggttctcgtcagatgattcactggttcaatcaggaataatcc 2099
DB 1201 CGCTTCATTTTGATGCTGCGGTTTCTGTCAGATGATTCACATGTTCAATACGGAATCC 1260
QY 2100 tgattctcaagctgcccacacagagctcgcagcaacagctagagcgaagcgcgtgg 2159
DB 1261 TGATTCTCAAGCTGCCCCAACAGAGGCTCGCAGCACCAAGCTAGAGCAGCAAGCGCGCTGG 1320
QY 2160 agatgacagtgcctgcagcgtgcagatgctcagaagcttttagaagcgccttagg 2219
DB 1321 AGATGACAGTGCCTGCTGACGCTGCGAGATGCTCAGAAGCTTTAGAACGCGCTCTAG 1380
QY 2220 taaagctgggcaacaaacagggcactactcaatgctttgggacagatcgcttctgctgctg 2279
DB 1381 TAAAGCTGGCAACAACAGGGCATACTCAATGCTTTTAGGACAGATCGCTTCTGCTGCT 1440
QY 2280 tgtgagcgcaggag-tectcccgcctgcagcaagttctatgg 2319
DB 1441 TGTGAGCGCAGGAGTTCTCCCGCTGCAGCAAGTTCTATAG 1481
```

```
RESULT 7
AX045135 1456 bp DNA PAT 24-NOV-2000
LOCUS
DEFINITION Sequence 5 from Patent WO0066739.
ACCESSION AX045135
VERSION AX045135.1 GI:11343734
KEYWORDS Chlamydomydia pneumoniae.
SOURCE
```

```
ORGANISM Chlamydomydia pneumoniae
REFERENCE 1 (bases 1 to 1456)
AUTHORS Murdin,A.D., Omen,R.P., Wang,J. and Dunn,P.
TITLE (Chlamydia) antigens and corresponding dna fragments and uses thereof
JOURNAL Patent: WO 0066739-A 5 09-NOV-2000;
Aventis Pasteur Limited (CA)
FEATURES
source
1..1456
Location/Qualifiers
/organism="Chlamydomydia pneumoniae"
/db_xref="taxon:83358"
101..>1452
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAG17379.1"
/db_xref="GI:11343735"
/translation="MWNPIGPGPIDTERTPPADLSAQGLESAANKSEAFORIGAE
AKPKESKTSVERVSLRSVAVNALMSLADKLGIASSNSSSSTSRSDVDTATATPT
PPPTFDYKTAQATYDTFTSTSLADIAALVSLQDAVTNFKDAADEETAIAAEW
ETKNDAVKVGAQITELAKVADNOAILDSGLKTSFDLLQALLQSVANNKRAELL
KEMQDNPVPGKTPAQAQSLVDOTDATQIEKDGNAI RDVAFAGNAGAVENAKSN
NSLSNDSAKAIAATTAQIAEQKFPSPILQAEQWVIOAEKDLKNIKPADGSDV
PNPFTVGGSKQGSIRVSMLLDDENETASILMSGFRQMHMFHTFNPDSQAA
QOELAAQARAARKAAGDDSAALADQAALKALEALGKAGQQQGLNALGQIASAAVVVSA
G"
BASE COUNT 452 a 308 c 331 g 365 t
ORIGIN
Query Match 56.9%; Score 1447; DB 6; Length 1456;
Best Local Similarity 99.7%; Pred. No. 3.8e-293;
Matches 1450; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 840 ataaaatttttaaaaaacaggtcgacattattattagtagagagctttttttttttttt 899
DB 1 ATAAATCTTTAAACACAGGCTCGCATTAATTATTAGTAGAGAGCTTTTTTTTATTTTTT 60
QY 900 ataataaactaaaagatttttttttttttttttttttttttttttttttttttttttttt 959
DB 61 ATATAAACCTAAAGATTTTTTATTATTTTGGATTTTATGTTTAAATCTATTGGTCC 120
QY 960 aggtctctatagcaaaacagacacacacctcccgagatctttctgctcaagattgga 1019
DB 121 AGGTCTATAGCAAAACAGAACACACACCTCCCGCAGATCTTCTGCTCAAGGATTGGA 180
QY 1020 ggcagtgacacaaataagagtcggaagctcaagaatagcaggtgcggaagctaaagc 1079
DB 181 GCGAGTGCACAAATAAGAGTGCAGGAGCTCAAGAGATAGCAGGTGCGGAAGCTAAGCC 240
QY 1080 taaagaatctaaagaccgattctgttagagcgtgagcattctgtcgttcgagtgaaatgc 1139
DB 241 TAAAGAACTTAAGACCGATTCTGTAGACGATGGAGCATCTTGGCTTCTGAGTGAATGC 300
QY 1140 tctcatgagctgcgagataagctggtgattcttcttagtaacagctcgtcttctactag 1199
DB 301 TCTCATGAGCTGCGCAGATAAGCTGGGTATTGCTTCTAGTAAACAGCTGCTTCTTACTAG 360
QY 1200 caagctcgcagagctggactcaacgacagcagccagcactacgcctccaccacagctc 1259
DB 361 CAGATCTGCAGAGTGGACTCAACGACAGCAGCAGCAGCTACGCCCTCCTCCACCCAGTT 420
QY 1260 tggatgattataagactcaagcgaacaaagcgttacgatactatttacctcaaacactact 1319
DB 421 TGATGATTATAAGACTCAAGCGCAACACAGCTTACGATACTTCTTTACTTCAACATCACT 480
QY 1320 agctcacatacagctgctttgtgagcctccagagatgctgctcactataataaagatac 1379
DB 481 AGCTGACATACAGGCTGCTTTGCTGAGCCTCCAGGATGCTGCTCACTAATAAAGATAC 540
QY 1380 agcggctactgatgaggaacacccgcaatcgctgcgagtgagggaactaagaatgcgcagtc 1439
```



[illegible]



On Jun 1, 2000 this sequence version replaced gi:7190933.

Db	COMMENT FEATURES	Location/Qualifiers
541	ATAGAGAAAGATGGAAATGCGATTAGGATGCATATTTTCAGGACAGACGCTAGTGGG	1..14698 /organism="Chlamydia muridarum" /db_xref="taxon:83560" /note="synonym: Chlamydia trachomatis MoPn" complement(105..1364) /gene="TC0906" complement(105..1364) /gene="TC0906" /note="conserved hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF39699.1" /db_xref="GI:7190934" /translation="MTWGLKQIGIFERSRVKQKPYHCVSRYHLPVLAAILPVG SWFLSEOWRYSYFLFPVYSSGLMFLAIGLREORLAAAGOLLETKIRKLTERDEGI KNIRETEKROKANRLKLNHDKLVERLQGAQREVYQAKGRYDHMEHMKLKHENKQ LQMLEAAVERNEKILENGELROELKETLAQOELHDEYOATVEQHNMKLDKQAYI GNLEKQVQDLMLCELRNLLOLEVGAKTELEPGRMSADVVVAOLVLEFRKIVRVEVTEA ADSLALRYTRTPDSAHNYSLACRQLFDDLRRENSLGMFLFIYAPFAQVFLPANALFKD THGLEDFLNKEDGVLEGFQAWERDLLESTERSGRKIVIKTKAFQAGTPFYCVVTL DKGFLAEHVLGVLYPAKASFFNTLSYI" 1869..2177 /gene="TC0907" 1869..2177 /gene="TC0907" /note="similar to GB:D10483 SP:P02378 GB:J01583 GB:M10428 GB:X04382; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="ribosomal protein S20" /protein_id="AAF39700.1" /db_xref="GI:7190935" /translation="MGEIMAPRPSKVKVGPQKPSAEKRVITSKKKQLRNQSFKSRVK TTLKFFELAVQSGDVASISAGLSVSIADKAVKRGIFRKGADRVKARASGACPTA " 2399..3202 /gene="TC0908" 2399..3202 /gene="TC0908" /note="conserved hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF39701.1" /db_xref="GI:7190936" /translation="MAAAVASTVPAGCLLSSSITALPRDTAALRSKILGYINDFLDL LSSSFALARLAPGASWASTADQQQSTFTEELTGVVHGANDAIQTGCLTOLLTGGHFFK TOAGSFQVDTSSNKRVLSPSLVSKVTRLSKMLGTVKFMSPLVPVQLGSHFALL LGLSLAFGTSSAFDAENTREVCAITLTKDKCCDSTGCSFMERFKRARGMARRG CSVFDLLAAQAFVCVETVSNAFMGVHTAFIVGIFCFLSALGNVLLIAAS" complement(3257..5884) /gene="TC0909" complement(3257..5884) /gene="TC0909" /note="conserved hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF39702.1" /db_xref="GI:7190937" /translation="MSSYILGNFRNTVSGGLFKIKLTNPCSDFKKRAKPAIDIELIN SGLYILRLSLVALEAGYLVGSVINPKNKIPFGGEGIVRRFQEGSVLSSRTISGNVTA DIKQSTAKTLVTITSDGLNTLIEDIPELPMQTQVGSISAALILMETYKSKPSLNETQ TVFGNAVYPSDESIIKAIKKEMEOEIOFKDLIIITAAQTAQASDOAIEESLRDEEF YVSEFNTFKVEDLWKYPATIGANTQEMLEKCSALGPDIPISLTLDNVKANAKSLVUK AWNEAFSNVLEVEPALGSGQIIIEAVLMKIPYATKTLITDEIHKIEQALSPPEVM DYLITQKQEGICKGAVKAFQATQNTQNVVRADVKENIEELEIKTKTSFQQAQASLESIM
1744	gctgtagaataatgctaaatctaataacagtaagcaacatagattcagctaaagcagca	1803
601	GCTGTGAGAAAATGCTAAATCTAATAACAGTATAGACACATAGATTTCAGCTAAAGCAGCA	660
1804	atcgctactgctaaagacacacaaatagctgaagctcagaaaaagttcccccagctctccaatt	1863
661	ATCGCTACTGCTAAAGACACAAATAGCTGAAGCTCAGAAAAAGTTCCCCGACTCTCCAATT	720
1864	cttcaagaagcgggaacaaatggtatatacagcgtgagaagaatcttaaaaaatacaaacct	1923
721	CTTCAAGAAGCGGACAAATAGTAAATACAGGCTGAGAAGATCTTAAAAATATCAAACT	780
1924	gcagatggtctctgatttccaaatccaggaaactacagttgaggtcccaagcaacaagga	1983
781	GCAGATGGTCTGATGTTCCAAATCCAGAACTACAGTTGGAGGCTCCAAGCAACAAGGA	840
1984	agtagtattggttagtattcgtgttcacatgctgttagatgctgtaaatgagacgcgt	2043
841	AGTAGTATTGTTAGTATTGCTGTTCCATGCTGTTAGATGATGCTGAAAAAGAGACCGCT	900
2044	tccattttatgctcgtggtttcgtcagatgattcacatgtttccaataaggaataatcctgat	2103
901	TCCATTTTGAATGCTGGGTTTCTGTCAGATGATTCACATGTTTCATACGGAATAATCCTGAT	960
2104	tctcaagctgccccaacagagagctcgcagcacagctagacagcagcgcgtgagagat	2163
961	TCTCAAGCTCCCCAACAGGAGCTCGCAGCACAAAGCTAGACAGCGCAAAAGCGCTGGAGAT	1020
2164	gacagtctcctcagcgcgtgcaaatgctcagaagctcttagaagcgcgtctaggtaaa	2223
1021	GACAGTCTGCTGCGCGCTGGCAGATGCTCAGAAGCTTTAGAACCGGCTCTAGGTAAA	1080
2224	gctgggcaacaacagggcactcaatgctttgggacagatgctctgctgctgtgtgtg	2283
1081	GCTGGCAACAACAGGCGACTACTCAATGTTTAGGACAGATCGCTTCTGCTGCTGTTGTG	1140
2284	agcgcagag-tcctccgcgtcagcaaatcttatg 2319	
1141	AGCGCAGAGTCTCTCCCGCTGCGCAAGTCTCTATAG 1177	
RESULT 9	AE002357/c	
LOCUS	AE002357 14698 bp DNA BCT 26-MAY-2000	
DEFINITION	Chlamydia muridarum, section 84 of 85 of the complete genome.	
ACCESSION	AE002357 AE002160	
VERSION	AE002357.2 GI:8163344	
KEYWORDS		
SOURCE	Chlamydia muridarum.	
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	
REFERENCE	1 (bases 1 to 14698)	
AUTHORS	Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.	
TITLE	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39	
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)	
MEDLINE	20150255	
PUBMED	10684935	
REFERENCE	2 (bases 1 to 14698)	
AUTHORS	Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA	







Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

1 (bases 1 to 10954)  
 Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,  
 Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
 Koonin, E.V. and Davis, R.W.  
 Genome sequence of an obligate intracellular pathogen of humans:  
 Chlamydia trachomatis  
 Science 282 (5389), 754-759 (1998)  
 9900809  
 9784136  
 2 (bases 1 to 10954)  
 Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hymen, R.W.,  
 Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
 Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
 Nat. Genet. 21 (4), 385-389 (1999)  
 9920606  
 10192388  
 3 (bases 1 to 10954)  
 Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,  
 Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
 Koonin, E.V. and Davis, R.W.  
 Direct Submission  
 Submitted (20-MAY-1998) Program in Infectious Diseases, University  
 of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA  
 Location/Qualifiers  
 1..10954  
 /organism="Chlamydia trachomatis"  
 /strain="D/UW-3/CX"  
 /db\_xref="taxon:813"  
 complement(146..2089)  
 /gene="CT622"  
 complement(146..2089)  
 /gene="CT622"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CHLPN 76kDa Homolog"  
 /protein\_id="AAC68226.1"  
 /db\_xref="GI:3329069"  
 /translation="MESGPESVSSNOSNMPLINGQIASNETKSTKESAPSPASS  
 SVSWSFSSAKHALISLDAILKNNSTPSLSOLEASTSVTVRAARDYNEAKS  
 FDTAKSGLNATIAEYETKMDALQMERAKOAEVTRKEALQKQEVLDK  
 LNQLVLEKQNTLKTETTTSDQIPAINSOLEINKNSADQIIKDLLEGONISTEAV  
 LTNAGEVTKASEAGIKLQALQSDVDQSOAAVLAQONNSPDNTAATKKLIDAA  
 ETKNELQKGTLDSPLVKKAEEQISQAQKIQEIKPSGSDIPVIGPSSGAASAGS  
 AVGALKSSNSGRIISILLDDVDNMAATAMQFRSMIEQFNVPNPATAKELQAEQAL  
 TMSQLVGADELPAEQAIKDALAQALQKOPSTDGLATAMGOVAFAAKVGGSAGT  
 AGTVQNNVKLTAFSTSSSSYAAALSDGYSAYKTLSLSYSSRSQVQSAISQTN  
 PALRSVSRSGIESQSRADASQRAAETIVRDSQTLGDVYSRLQVLDLSMTIIVSNPQ  
 VNQEEIMQKLTAISKAFQGYPAVQNSADSLQKFAAQLEFEFVDGERSLAESRENAF  
 RKQAPFIOQVLNIALSFGYLS"  
 complement(2197..3495)  
 /gene="CT623"  
 complement(2197..3495)  
 /gene="CT623"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CHLPN 76kDa Homolog"  
 /protein\_id="AAC68227.2"  
 /db\_xref="GI:6578109"  
 /translation="MKKFIKGFVGAALLACGSTNLAFAQASMSQSLWSVEDLSYL  
 SSKGFVETRRQGLVRLAGDVRAARIYAKEDLETTQTPAKMPLNRYSEFNLYVDY  
 TANSWNTSKMNVFTIAGGESSAAGLDINRAFLGYRFYKNPETQAEVFAEIGRSLGD  
 IPDSYQFNSFDGIHYAARISEKLPITMIVHGGPFVYVNAEYEAIVVVAIINKL  
 PGNVVKTQSVNDNTLTAKNDPADASAAQAPKNTKYDLVWQVLVKGSTAMPFNG  
 QTNKLYTGALNPLAEIPENKQSTTPTTKTNGKENHAFIGCSLGGVRRAGDWS  
 ATVRYETCALAIPEDVAGIRGNOMKWFQAQIKQGLDPKESNGFTNYKGVSYQV  
 MGLTDSVFRAYAAYSKPNADNGISDFTTKYKLDGLISF"  
 complement(3738..5348)  
 /gene="mvln"  
 /note="CT624"  
 complement(3738..5348)  
 /gene="mvln"  
 /codon\_start=1

/transl\_table=11  
 /product="Integral Membrane Protein"  
 /protein\_id="AAC68228.1"  
 /db\_xref="GI:3329071"  
 /translation="MRKDDGSLVRSLENLNLGTFSSLTGMLREIVNATVFGADPLV  
 ASFWLAFTIFFRLKLLGPIGLAFIPHEFLRAONISRAATFFFRSFRFCYSAIL  
 FFLITELGVCWCVSGTSLFTLLTLLTLLPSGIFLMYTVNLTLLHCEKFFSVGL  
 APSVNVSWIGTVELARNYDPRNIFGLAVLVIGFLEWAVTLPGVWKFQGSKEVP  
 QERDSIRALIAPLSGLISMGIFOLNLLCDMLARYINEVGPLYLAYSVRIQOLGVRL  
 FGLGVFTVLLPAISRCVQDQEQHQQGDDLLRLSLKLTAVAVVYVWMLLILFALPGVRL  
 YHGVFPFTAVHAIIVEILRGYSIGSIIPMALAPLSALFYARRNYKVPMLVGLIAAVN  
 MYLNVIGCVICQVAVATYATSLVSGOLAMLYCAGKSLPTYKGLMRTFKESKTV  
 ITTILAAVITIGVNIYVTVYVPEITVTPKPLVSLFDQCGVFAESALFSLVLF  
 LAKLLKTDLMNLISFYQWKHQSLRN"  
 5521..6387  
 /gene="nfo"  
 /note="CT625"  
 5521..6387  
 /gene="nfo"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Endonuclease IV"  
 /protein\_id="AAC68229.1"  
 /db\_xref="GI:3329072"  
 /translation="MFILPPPQALLGAHTSAAGLHNLALYEGRDIGATTVQLFTANQ  
 ROKRRTLTQEMVDPQAFRIALNETSLYIMSHAGYLNPGAPNPPPEILTRVCMHQEIA  
 DCISLGSFVNRHPGAALSDSKESCLDRTIASFSQMAPLPENNPLVLTETAGGGS  
 LIGSSPEELAYLIQIKAHPIGVCLDTCHIFAAQYDSSVAGWEOVLKHFDAVIGLS  
 FLRAIHINDSVFALGKNKDEHAPIGEGCIGSDSCFCLMDQERTMLPKYLETPGGPD  
 WTKELRYLQKVC"  
 complement(6484..7113)  
 /gene="rs4"  
 /note="CT626"  
 complement(6484..7113)  
 /gene="rs4"  
 /codon\_start=1  
 /transl\_table=11  
 /product="S4 Ribosomal Protein"  
 /protein\_id="AAC68230.1"  
 /db\_xref="GI:3329073"  
 /translation="MARYCGPKNRIARRFGANIFGRNPLLRPNPPOGHGMRKK  
 SDVQLQEEQKQKACYLEQKLVKAYEVNKNQVQAEKLEKPECLDNIYVRL  
 GFAPKIFAAQOLYSHGVNKGKVDERSFVPGMQLSLKESKRLAIVTESLENKD  
 QSSLPALISLQKAKAFKELVVAPELDQIASQLPLPVNVSVICELSHRT"  
 7497..8480  
 /gene="ycaA"  
 /note="CT627"  
 7497..8480  
 /gene="ycaA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAC68231.1"  
 /db\_xref="GI:3329074"  
 /translation="MEKNYALAYYFGPVSNPHEALHKLQFKTMDVSGRIYISEE  
 GINGQFSQYDAERYANWLKQRPDFAIKFKIHIEENIFPRVTVKRYKVALGCS  
 VDTTKQKHISPEEHHEKLEQENCLVDVRNNYEWKIGHFENAVLPDIEFREPDA  
 DRLAKHDHPARTPYMCTGIRCELYSALLERGFKEVOLDGGVITAYGLKMTGKW  
 RKGLFVDFDRNAMPIDEADPNVSPITARCSLCSDTSDTYINCANTDNNLFTCCSCIA  
 THKGCCSECSQAIPRAFAERGNKPRRKHLCPTIEQSCCLKEQENQPA"  
 complement(8552..9427)  
 /gene="lspa"  
 /note="CT628"  
 complement(8552..9427)  
 /gene="lspa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Geranylgeranyl pyrophosphate synthase"  
 /protein\_id="AAC68232.1"  
 /db\_xref="GI:3329075"  
 /translation="MDYFSVYRSKVEKKLRSLRSLDGFQNGGLRDPTEYALLGGKRV  
 RPLVLGLFAEGVHKERVDLTAIEYIHTLTADLLPCMDDDDDMRKGFVSHKAFD  
 BASALLASVALIPAYAIRKNAKAKAVVCEQDVEAYEDILELIELRFGVGGILG



Qy	613	aagtttgggtccccaagaatgtctgctaactatgatcctaaagagcctaagtgttt	672
Db	2412	AAATATGTGTTTCTCAAGCTATAACAAGAGATTGGATTCCTTAAAGAATCTAACGGGTTT	2353
Qy	673	acaaattaaagatcttcgcctctatatgtatggcatacagatctctatcattc	732
Db	2352	ACTAACTATAAAGAGATTTCCTATCAGTTGTATGGGTCTGACAGATTTCGGTTTCTTTC	2293
Qy	733	agagcttatggggcttactccaaaccagcaaacgataaactgcgcagtgattttacttcc	799
Db	2292	CGAGCTTATGCTGTTATTCTAAGCGCTGCTACAGATAACCTTGTGTAGGACTTCACCTAT	2233
Qy	793	cgaagtttgatcagtgataatttcagcgtttttaagtcacaaattttaataaaactcttaa	852
Db	2232	CGTAAGTATGACCTAGGTTTAATTTCTTCATTCTAATCC - CTTAAGGGATGAGTCTTTA	2175
Qy	853	aaacaggcctgcattaattattagtgagagcgtttttttttttttttataataaaactaa	912
Db	2174	AGAGCCACACCCAAAATAACGGATGTGGCTTTTTTTTATTGCGCTTTTATAATTAAAGTAA	2115
Qy	913	aagatcttttattatttttttgagttttt	939
Db	2114	GGTTTTTTTATTTTTTTTATTGTTAT	2088
RESULT	11		
CTU50732			
LOCUS	CTU50732	3455 bp	DNA
DEFINITION	Chlamydia trachomatis mvin homolog, lorf2; possible membrane-protein, and 76 kDa protein homolog genes, complete cds.		
ACCESSION	U50732		
VERSION	U50732.1	GI:4165180	
KEYWORDS			
SOURCE	Chlamydia trachomatis.		
ORGANISM	Chlamydia trachomatis.		

```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCES
AUTHORS
JOURNAL
REFERENCE
AUTHORS
JOURNAL
REMARK
COMMENT
FEATURES
source

1 (bases 1 to 3455)
Myers,G.S., Grinvalds,R., Booth,S., Hutton,S.I., Binks,M.,
Kemp,D.J. and Sriprakash,K.S.
Expression of two novel proteins in Chlamydia trachomatis during
natural infection
Microb. Pathog. 29 (2), 63-72 (2000)
20368578
10906261
2 (bases 1 to 3049)
Myers,G.S.A. and Sriprakash,K.S.
Direct Submission
Submitted (05-MAR-1996) Molecular Genetics Unit, Menzies School
of Health Research, Rocklands Drive, Darwin, NT 0810, Australia
3 (bases 1 to 3455)
Sriprakash,K.S.
Direct Submission
Submitted (21-JAN-1999) Molecular Genetics, Menzies School of
Health Research, PO Box 41096, Casuarina, NT 0811, Australia
Sequence update by submitter
On Jan 21, 1999 this sequence version replaced gi:1255183.
Location/Qualifiers
1..3455
/organism="Chlamydia trachomatis"
/db_xref="taxon:813"
/note="serovar: B"
168..1778
/note="lorfl; likely cytoplasmic membrane protein;
possible virulence protein"
/codon_start=1
/transl_table=11
/product="mviN homolog"
/protein_id="AAD08715.1"
/db_xref="GI:1255184"
/translation="MRKDDDECLVRSFLNLLSGFTFSRLTGMLREIVMATYFQGA
ASFWLAFTIFFRLKILGGPILGLAFIPHFEEFLRAQNISRTATFFRSFRCFYC
FTLLTEILGLCYMCSCVTCGSLFDTLTLTILLPSCGIFILMVTNVTSLHCEKFKVF

```

APSVVNVSWIGTVFLARNYDPRNRIFGLAVVLVIGFILEWAVTLPGVMKFLGQSS



QERDSIRALHAPLISGILLSSMGIFQOLNLLCDMWLARYINEVGPPLYLMYSVRIQOLPVLH  
FGLGVFVLPALSRQVODHQOQGYDLRLFLKLVAVVMVMTGLFVFLPGLVRL  
YEHCVPEKTAHVAIVEIRGYSIIPMALPLVSAFYARRNYKVPMLVIGIIAAVYN  
MYLVIGICVKQAVLAVATSLVSGOLAMWYACAGKSLPTVYKGLMRTFFRESGTV  
ITLILAAVITIGVNIHTHYVFIPLVPTKPLVSLFDQGVFFAESALFLSVLFG  
LAKLLKTEDLMLNISFOYWKHOSILRN  
2022..2570  
CDS  
/note="lorf2; possible membrane-bound protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AAD08716.1"  
/db\_xref="GI:1255185"  
/translation="MKKIFYKGFVGVALLACGNTNIAFAQASMSQLSVDELDLYL  
SSKGFVTRRDGVLRARWIAKEDLETTQTPAKPLMPTNRYSEFNLYDY  
TAANSWTKMKNWYIAGGESSAGLDINRAFLGYRKNPQAEVFAEIGRSLGD  
IFDSTFSLIVISTEFIVLTRY"  
2688..3311  
CDS  
/note="lorf3; possible inner membrane protein; similar to  
C. pneumoniae 76 kDa protein; Method: conceptual  
translation supplied by author"  
/codon\_start=1  
/transl\_table=1  
/product="76 kDa protein homolog"  
/protein\_id="AAD08717.1"  
/db\_xref="GI:4165181"  
/translation="MKTSVDVNTLTAKTNDPADASAAQPAKPNKYDYLVWQVLVCK  
STAMPFNGOKNLYTGAYLFNLAEPENWKNQSTTPATKTINGKENHAWFGCSUG  
GVRRAGWDSALVRYEYKLRPEIDVAGIGRNQKQYFQAIRKGLDPKESNGFTNYK  
GVSYQVWGLTDSVFRAYAAYSKPANDNLGSDTYRKYDLGLISF"  
BASE COUNT 928 a 502 c 780 g 1145 t  
ORIGIN

Query Match 8.6%; Score 217.6; DB 1; Length 3455;  
Best Local Similarity 59.5%; Pred. No. 1.9e-35;  
Matches 469; Conservative 0; Mismatches 269; Indels 50; Gaps 4;  
QY 194 cttgggttagaaggattccatcggtttgcttaaacagttttttggaatagtg 253  
DB 2641 CTTGGGTCGTGGGAAGTATTGTAATAAACTCCAGGAATTCGTGTGGAACGAGTG 2700  
QY 254 ttgtcagctggaacacattcgttcc-----ttcagaacccctccactacagaaa 301  
DB 2701 TTGTTGACTGGATAGCTTAACAGCAAAACGAATGATCCAGCAGCGAAGCGTGCAC 2760  
QY 302 aagctgctacaaacgtctgaatacacaatactgtgttggcagtggtcgtcggaagac 361  
DB 2761 AACGAGCTAAACCTAATACCAAGTACGATTATTAGTAGGCAATGGTTGGTGGGAAG 2820  
QY 362 atagtcaggttccttgatcaatgagacagaaaagcctctatatctttatgagctttct 421  
DB 2821 GCACAGTATGCCATGCTTTAATGGCAACAAACAAAATCTTACACTTACGGAGCCTATC 2880  
QY 422 taatgaaccttttag-----caaggctcaggaaga 451  
DB 2881 TCITTAATCCATTAGCGGAATACCAGAACTGGAACAAATCAACAACTCTGCAACCA 2940  
QY 452 ctacgttaaatgaaagaaacactagcttggtttattggaggaacttttaggggaactca 511  
DB 2941 AAATTACAAATGTAAGGAAACCATGCTTGGTTCATCGGCTGTTCTAGGCGGGTTC 3000  
QY 512 gaaagctgagactggtctgcacacagtcattagatgattgcgaagcctttgctcagttc 571  
DB 3001 GACAGCTGGAGACTGGTCTGCATAGTTCGTTATGATATGTAA-----CCTTCGAC 3054  
QY 572 cagaatagattgtttcaggattggccggtggtattattattaaagtttgggttcgccaag 631  
DB 3055 CAGAAATTGATGTCGCGGGTATTGGTCGCGGAAACCAAAATGAATATGTTGTTGCTCAAG 3114  
QY 632 caattgctgaactatgactcaaaagagcctaagttttacaattataaaggatttt 691  
DB 3115 CTATCAAAACAGGATTGGTCCCTAAGAACTAACGGCTTTTACTAATATAAGGAGTTT 3174

QY 692 ccgctctatatgtatggcatcacagattctctatcattcagagcttatggggtctact 751  
DB 3175 CCTATCAGTTGTTATGGGCTGACAGATTCGGTTCTTCCGAGCTTATGCTGCTATT 3234  
QY 752 ccaaacacagcaaacgataactcgcagtgattttacttctccgaaagtgtgactagta 811  
DB 3235 CTAAAGCTGCTAACGATAACCTTGGTAGGACTTCACTATCGTAAGTATGACCTAGTT 3294  
QY 812 taatttcagcgttttaagtcacaaatttaataaaatcttttaaaacaggtcgcattatt 871  
DB 3295 TAATTCTTCATTCTAATCC--CTTAAGGGATAGTCTTTTAAGAGCAACACCAAAATA 3352  
QY 872 attaggaagcttttttttttttttttttttaataaaactaaaagattttattattttt 931  
DB 3353 ACGGCTGGGGCTTTTATTCCTTTTATTAATAAAGTAAGTGTCTTCTATTATTTTAA 3412  
QY 932 gaggttttt 939  
DB 3413 TATTGTAT 3420  
RESULT 12  
AY029367 6849 bp DNA SYN 09-MAY-2001  
LOCUS Synthetic construct his-tagged human friend leukemia integration 1  
DEFINITION transcription factor (FLI1) gene, complete cds.  
ACCESSION AY029367  
VERSION AY029367.1 GI:14017400  
KEYWORDS synthetic construct.  
SOURCE artificial construct.  
ORGANISM Uhbi.B.T.S.; Rainey,D.R. and Meredith,D.M.  
REFERENCE 1 (bases 1 to 6849)  
AUTHORS Direct Submission  
TITLE Submitted (09-APR-2001) Molecular Medicine Unit, Leeds University,  
JOURNAL Beckett St, Leeds LS9 7TF, United Kingdom  
FEATURES  
Location/Qualifiers  
1..6849  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/chromosome="11"  
/clone="DH6"  
/note="cloned into pCDNA3.1 myc his(-) expression vector;  
derived from U937"  
959..2398  
/gene="FLI1"  
959..2398  
/gene="FLI1"  
/note="E1S transcription factor; involved in the  
pathogenesis of Ewings sarcoma following translocation  
with EWS gene on chromosome 22; may have oncogenic  
potential when fused with EWS gene"  
/codon\_start=1  
/transl\_table=11  
/product="his-tagged human friend leukemia integration 1  
transcription factor"  
/protein\_id="AAK50442.1"  
/db\_xref="GI:14017401"  
/translation="MDGTIKEALSVSDQSLFSDAYGAAHLPKADMTASGSPDYGO  
PHKINPLPQEWENINQVRNVKREYDHMGSRSPVDCSVSKSLVGGESNPMNY  
NSYMDKNGPPPPNMTINERRVYVADPTLTWQHRVQWLEWAKKEYSLMEIDTSPFQ  
NMDGKELCKMKNKDLRATTLTYNTEVLSLHSLYDRESLLAYNTTSHDQSSRLSVRE  
DPYSDSVRRGAWNMNGLNKSPLGGAQTISKTQRPOPDYQILGPTVSSSLANP  
GSGOIQWQFLLELLSDSANASCITWETNGEFMTDFDEVARWGRKSKPNNDK  
LSRALRYVDKNIWTKVHGKRYAYKDFHGAQALQPHPTESMYKYPDSIYSPSYH  
AHQOKVNEPVPSPMPVTSSFFGAASQYVTSPTGGIYPNPVPRHPNTHVPSHLGS  
YVYKLGFEQKLISEEDLNSAVDHHHHH"  
BASE COUNT 1632 a 1844 c 1734 g 1639 t  
ORIGIN  
Query Match 7.9%; Score 200.2; DB 12; Length 6849;











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:48:06 ; Search time 685.35 Seconds  
(without alignments)  
3183.618 Million cell updates/sec

Title: US-09-391-606-1

Perfect score: 2545

Sequence: 1 ttgcggtgctgtaacggtg.....tgccactccactgtccttt 2545

Scoring table: IDENTITY\_NUC

Gap 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N\_Geneseq\_1101:\*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:\*
  - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:\*
  - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:\*
  - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:\*
  - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:\*
  - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:\*
  - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:\*
  - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:\*
  - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:\*
  - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:\*
  - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:\*
  - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:\*
  - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:\*
  - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:\*
  - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:\*
  - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:\*
  - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:\*
  - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:\*
  - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:\*
  - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:\*
  - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:\*
  - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2225.2	87.4	2238	21	AAD02066
2	2111.2	83.0	1230025	20	AA91990
3	1459.4	57.3	2156	21	AAD02063
4	1447	56.9	1456	21	AAD02065
5	1155.4	45.4	1852	21	AAD02064
6	832.8	32.7	1550	21	AA28411
7	179	7.0	1909	21	AAA27110
8	171	6.7	4912	20	AA08423
9	171	6.7	4912	21	AAA71402
10	171	6.7	5108	22	AAC86258
11	171	6.7	5108	22	AAC86259

12	171	6.7	5111	22	AAC86262	CMV IE promoter eh
13	171	6.7	5185	22	AAC86254	PGHRH-4 construct.
14	171	6.7	5188	22	AAC86266	Plasmid GHRH1-44YW
15	171	6.7	5254	22	AAC86287	Plasmid PGHRH1-44W
16	171	6.7	7285	20	AA89797	DNA of pVR 1012-GP
17	171	6.7	7285	20	AA89797	Plasmid pVR 1012-GP
18	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
19	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
20	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
21	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
22	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
23	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
24	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
25	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
26	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
27	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
28	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
29	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
30	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
31	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
32	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
33	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
34	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
35	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
36	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
37	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
38	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
39	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
40	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
41	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
42	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
43	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
44	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G
45	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-G

ALIGNMENTS

RESULT	1
AAD02066	
ID	AAD02066 standard; DNA; 2238 BP.
XX	
AC	AAD02066;
DT	26-MAR-2001 (first entry)
DE	C. pneumoniae 76 kDa protein truncation mutant fusion gene.

KW	76 kDa protein; bactericidal; diagnosis; prevention;
KW	pneumonia; upper respiratory tract disease; sinusitis;
KW	acute respiratory disease; cough; sore throat; hoarseness; fever;
KW	vaccine; immunisation; treatment; truncation mutant; fusion gene;
KW	ds.
OS	Chlamydia pneumoniae.
OS	Synthetic.
XX	
FT	Key
FT	misc_feature
FT	Location/Qualifiers
FT	1..665
FT	/*tag= a
FT	/note= "This part of the sequence is unrelated to
FT	C. pneumoniae 76 kDa gene"
FT	766..2238
FT	/*tag= b
FT	/product= "Truncated Chlamydia pneumoniae
FT	76kDa protein"
FT	2122..2238
FT	/*tag= c
FT	/note= "This part of the sequence is unrelated to
FT	C. pneumoniae 76 kDa gene"
PN	W0200066739-A2.



XX	09-NOV-2000.
PD	
XX	
XX	03-MAY-2000; 2000WO-CA00511.
Pf	
XX	
XX	03-MAY-1999; 99US-0132270.
PR	
PR	30-JUN-1999; 99US-0141276.
XX	
XX	(AVET ) AVENTIS PASTEUR LTD.
PA	
XX	Murdin AD, Oomen RP, Wang J, Dunn P;
PI	
XX	WPI: 2000-687542/67.
DR	P-PSDB; AAU71957.
XX	
PT	Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
PT	useful for vaccinating against Chlamydia infections -
XX	
PS	Claim 32; Fig 3; 112pp; English.
XX	
CC	The present sequence is a DNA coding for a fusion protein comprising a
CC	truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded
CC	residues. C. pneumoniae 76 kDa protein is used in the
CC	diagnosis, prevention and treatment of C. pneumoniae
CC	infections (e.g. pneumonia, upper respiratory
CC	tract disease, bronchitis, sinusitis and acute respiratory
CC	disease such as cough, sore throat, hoarseness, fever; and
CC	abnormal chest sounds on auscultation). C. pneumoniae sequence
CC	is also used as vaccines for immunising humans against diseases
CC	caused by C. pneumoniae.
XX	
SQ	Sequence 2238 BP; 689 A; 459 C; 497 G; 593 T; 0 other;
Query Match 87.4%; Score 2225.2; DB 21; Length 2238;	
Best Local Similarity 99.6%; Pred. No. 0;	
Matches 2230; Conservative 0; Mismatches 8; Indels 0; Gaps	
QY	175 atgacaaaaaacattatgcgttgggttagaaggagttctcaatcggttgccctaacaagc 234
Db	
Db	1 atgacaaaaaacattatgcgttgggttagaaggagttctcaatcggttgccctaacaagc 60
QY	235 ttttttgaaatgagtgttcgtactggaaacaaattcgttccttcgaagaacctcccact 294
Db	
Db	61 ttttttgaaatgagtgttcgtactggaaacaaattcgttccttcgaagaacctcccact 120
QY	295 acagaaaagctgctcaaacgctatgaaatacaaatctgtgttgccagtggtcgtc 354
Db	
Db	121 acagaaaagctgctcaaacgctatgaaatacaaatctgtgttgccagtggtcgtc 180
QY	355 ggaagcatgtagtgcgttccctggatgcataatggacagaaaaagcctctatatctttatgga 414
Db	
Db	181 ggaagcatgtagtgcgttccctggatgcataatggacagaaaaagcctctatatctttatgga 240
QY	415 gctttcttaatgaaccctttagcaaggctacgaagactacgtttaattggaagaagaac 474
Db	
Db	241 gctttcttaatgaaccctttagcaaggctacgaagactacgtttaattggaagaagaac 300
QY	475 ctagcttggtttattgaggaaaccttaggggggactcagaaaagctggagactggtctgcc 534
Db	
Db	301 ctagcttggtttattgaggaaaccttaggggggactcagaaaagctggagactggtctgcc 360
QY	535 acagtaagttatgagtatgtcgaaagccttgtcagttccagaaatagatatgtttcaaggatt 594
Db	
Db	361 acagtaagttatgagtatgtcgaaagccttgtcagttccagaaatagatatgtttcaaggatt 420
QY	595 ggccgttggttaatttataaagttttggttcgcccaagcaattgctgctaactatgatcct 654
Db	
Db	421 ggccgttggttaatttataaagttttggttcgcccaagcaattgctgctaactatgatcct 480
QY	655 aaagaggtctaagtttttaacaaattataaaggatttttcgcgtctatatatgtatggcatc 714
Db	
Db	481 aaagaggtctaagtttttaacaaattataaaggatttttcgcgtctatatatgtatggcatc 540



Qy 1795 aaagcagcaatcgtactgctaagaacacaaatagctgaagctcagaaaaagttcccccac 1854  
 Db 1621 aaagcagcaatcgtactgctaagaacacaaatagctgaagctcagaaaaagttcccccac 1680  
 Qy 1855 tctcaaatcttcaagaagcggaacaaatgtaatacacaggctgagaaagatcttataaaat 1914  
 Db 1681 tctcaaatcttcaagaagcggaacaaatgtaatacacaggctgagaaagatcttataaaat 1740  
 Qy 1915 atcaaacctcagatggttctgatttccaaatccagaaactacagtggagggctcccaag 1974  
 Db 1741 atcaaacctcagatggttctgatttccaaatccagaaactacagtggagggctcccaag 1800  
 Qy 1975 caacaagaagtagtattgtagtattcgtgttccatgctggttagatgagctgaaaaat 2034  
 Db 1801 caacaagaagtagtattgtagtattcgtgttccatgctggttagatgagctgaaaaat 1860  
 Qy 2035 gagaccgctccattttagtctggttctggttccagatgattcaacatttcaatacggaa 2094  
 Db 1861 gagaccgctccattttagtctggttctggttccagatgattcaacatttcaatacggaa 1920  
 Qy 2095 aatcctgatttcaagctgccaaacagagagctcgagcacaagctagagcagcgaagccc 2154  
 Db 1921 aatcctgatttcaagctgccaaacagagagctcgagcacaagctagagcagcgaagccc 1980  
 Qy 2155 gctgagatgacagtgtcgtgcagcgtgcagatgtctcagaaagctttagaagcggct 2214  
 Db 1981 gctgagatgacagtgtcgtgcagcgtgcagatgtctcagaaagctttagaagcggct 2040  
 Qy 2215 ctagctaaagctggcgaacacagagcactacatcaatgcttgggacagactcgtctgct 2274  
 Db 2041 ctagctaaagctggcgaacacagagcactacatcaatgcttgggacagactcgtctgct 2100  
 Qy 2275 gctgttggagcagagagctcccgctgcagcagagttctatggtatcgagctcggtac 2334  
 Db 2101 gctgttggagcagagagctcccgctgcagcagagttctatggtatcgagctcggtac 2160  
 Qy 2335 caagcttcagtagaacaacaaactatctcagaagagatctgaatagcgcgtcgaccat 2394  
 Db 2161 caagcttcagtagaacaacaaactatctcagaagagatctgaatagcgcgtcgaccat 2220  
 Qy 2395 catcatcatcatctga 2412  
 Db 2221 catcatcatcatctga 2238

RESULT 2  
 AAX91990/c  
 ID AAX91990 standard; DNA; 1230025 BP.  
 XX  
 AC AAX91990;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
 DE  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; ss.  
 OS  
 XX Chlamydia pneumoniae.  
 PN  
 XX WO9927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-1B01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffiths R;

XX WPI; 1999-357842/30.  
 DR  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 PS  
 PS Claim 1; Page 291-611; 1912pp; English.  
 XX  
 CC The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AAV34584-V35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AAV34584-V35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.  
 XX  
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 83.0%; Score 2111.2; DB 20; Length 1230025;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 2138; Conservative 0; Mismatches 8; Indels 2; Gaps 2;  
 Qy 174 catgacaaaaaacattatgcttgggttagaaggattctcaatcgtttgcctaaaca 233  
 Db 829364 CATGACAAAAAACATTATGCTTGGTGTAGAGGGATTCTCAATCGTTGCCCTAAACA 829305  
 Qy 234 gtttttggaaatgtagtggctgactggaaacattcgtctcctcagaacacctcac 293  
 Db 829304 GTTTTTTGTGAATAGTGTGTGCTGGAACACATTCGTTCTTTCAGAAACCTCCAC 829245  
 Qy 294 lacgaaaaagctctacaacgcctatgaatacaaaactgctgtttggcagctgctgct 353  
 Db 829244 TACAGAAAAAGCTGCTACAAACGCTATGAATACAAATCTGTGTGGCAGTGGCTCGT 829185  
 Qy 354 cggaaagcatagtcaggttctcttggatcaatggacagaaaaagcctctatcttctatgg 413  
 Db 829184 CGGAAGCATAGTCAGGTTCCTTGGATCAATGGACAGAAAAAGCCTCTATATCTTTATGG 829125  
 Qy 414 agctttttaaatagaaccttttagcaaaagctacgaagactacgttaaatggaaagaaaa 473  
 Db 829124 AGCTTTCTTAATGAACCCCTTTAGCAAAAGCTACGAAGACTACGTAAATGGAAGAAAA 829065  
 Qy 474 cctagcttgggttattggaggaacttttaggggactcagaaagctggagactgctgctgc 533  
 Db 829064 CCTAGCTTGGTTATTGGAGGAACTTTAGGGGACTCAGAAAAAGCTGGAGACTGGTCTGC 829005  
 Qy 534 cacagtacgttatgagtagtgcgaagccttgcagttccagaaatagatgtttcagggat 593  
 Db 829004 CACAGTACGTTATGAGTATGTCGAAGCCTTGTGGTTCAGAAATAGATGTTTTCAGGGAT 828945  
 Qy 594 tggcgttggttaattattaaagtgttgggtcccaagaactgtgctgaactatgatcc 653  
 Db 828944 TGGCGCTGGTAATTTATTAAAGTTTGGTTCGCCCAAGCAATTTGCTGAATATGATGCC 828885  
 Qy 654 lazagaggctaatagttttacaataataaaggattttccgctctctatctatgtatggcat 713  
 Db 828884 TAAAGAGGCTAATGCTTTTACAAATATAAAGGATTTTCCGCTCTATATATGATGGCAT 828825  
 Qy 714 cacagattctctatcattcagagcttaagggtcttactccaacacagcaacgataaaact 773  
 Db 828824 CACAGATTCTCTTATTCATTCAGAGCTTATGGGGCTTACTCCAAACAGCAACGATAAACT 828765  
 Qy 774 cggcagtgatttttacttccgaaagttagctagatataaattcagcgtttttaagctcaa 833  
 Db 828764 CGGCAGTGATTTTACTTTCCGAAAGTTTGTAGTATGATTAATTTTCAGCGTTTTTAAGTCAA 828705  
 Qy 834 attttaataaaatctttaaaacagcgcgcattattattagtagagacttttttta 893  
 Db 828704 ATTTTAATAAATCTTTAAACACAGGCTCGCATTAATATTATTAGTAGAGCTTTTTTTTA 828645



Db	827564	GCACCAAGGAAGTAGTATTGGTAGTATTCGTGTTCCATGCTGTAGATGATGCTCAAAA	827505
Qy	2034	tgagaccgctccattttgatctgtctgggttttqtcagatgattccaatcacgga	2093
Db	827504	TGAGACCGCTCCATTTTGATGCTCTGGGTTTCGTACAGATGATTACATGTTCAATACGGA	827445
Qy	2094	aatcctgattctcaagctgcccacaacagagagctcgcagcacagcttagcagcgcgaaagc	2153
Db	827444	AAATCCTGATTCTCAAGCTGCCCAACAGAGAGCTCGCAGCACAAAGCTAGACGACGGAAGC	827385
Qy	2154	cgctggagatgacagtgctgctcagcgcgtgagatgctcagaaagctttagaagcggc	2213
Db	827384	CGCTGGAGATGACAGTCTCTCTCGACGCTGGCGAGATGCTCAGAAAAGCTTTTGAAGCGGC	827325
Qy	2214	tctaggtaaagctggggccaacacagggcatactcaatgcttttgggacagatcgctcttcg	2273
Db	827324	TCTAGGTAAAGCTGGGGCAACACAGGGCATACTCAATGCTTTAGGACACATCGCTTCTGC	827265
Qy	2274	tgctg-ttgtgagcgcaggag-tcttcccgctgcagcaagttctatgg	2319
Db	827264	TGCTGTTGTGAGCGCAGGAGTTCTTCCCGCTGCAGCAAGTTCTATAG	827217
RESULT 3			
AAD02063			
ID AAD02063 standard; DNA; 2156 BP.			
XX	AC	AAD02063;	
XX	AC	AAD02063;	
DT	DT	26-MAR-2001 (first entry)	
XX	DE	Chlamydia pneumoniae 76 kDa full-length protein encoding DNA.	
XX	DE	76 kDa protein; bactericidal; diagnosis; prevention; treatment;	
KW	KW	pneumonia; upper respiratory tract disease; bronchitis; sinusitis;	
KW	KW	acute respiratory disease; cough; sore throat; hoarseness; fever;	
KW	KW	vaccine; immunisation; ds.	
XX	XX	Chlamydia pneumoniae.	
OS	OS	Key	Location/Qualifiers
PH	PH	CDS	101..2056
FT	FT		/*tag= a
FT	FT		/product= "Chlamydia pneumoniae 76kDa protein"
XX	XX	W0200066739-A2.	
XX	XX	09-NOV-2000.	
XX	XX	03-MAY-2000; 2000WO-CA00511.	
XX	XX	03-MAY-1999; 99US-0132270.	
PR	PR	30-JUN-1999; 99US-0141276.	
XX	XX	(AVET ) AVENTIS PASTEUR LTD.	
XX	XX	Murdin AD, Oomen RP, Wang J, Dunn P;	
XX	XX	WPI; 2000-687542/67.	
DR	DR	P-PSDS; AAY71954.	
XX	XX	Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,	
PT	PT	useful for vaccinating against Chlamydia infections -	
XX	XX	Claim 2a; Fig 1; 112pp; English.	
XX	XX	The present sequence is a DNA coding for Chlamydia pneumoniae	
CC	CC	76 kDa protein. C. pneumoniae 76 kDa protein	
CC	CC	is used in the diagnosis, prevention and treatment	
CC	CC	C. pneumoniae infections (e.g. pneumonia, upper respiratory	
CC	CC	tract disease, bronchitis, sinusitis and acute respiratory	
CC	CC	disease such as cough, sore throat, hoarseness, fever; an	



CC abnormal chest sounds on auscultation). C. pneumoniae sequence  
CC is also used as vaccines for immunising humans against diseases  
CC caused by C. pneumoniae.

SQ Sequence 2156 BP; 672 A; 461 C; 471 G; 552 T; 0 other;

Query Match 57.3%; Score 1459.4; DB 21; Length 2156;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1474; Conservative 0; Mismatches 6; Indels 1;

1:

[illegible]

D	b	901	tggagctgtagaaaa	tgctaaatctataacacgataagcaacatagattcagctaaagc	960
Q	y	1800	agcaatcgctactgctaagacacaaatagctgaagctcagaaaaagtcccccga	ctctcc	1859
D	b	961	agcaatcgctactgctaagacacaaatagctgaagctcagaaaaagtcccccga	ctctcc	1020
Q	y	1860	aattctcaagaagcggaaacaatggtaatacagagctgagaaaaatcttaaa	aatatcaa	1919
D	b	1021	aattctcaagaagcggaaacaatggtaatacagagctgagaaaaatcttaaa	aatatcaa	1080
Q	y	1920	acctgcagatggtctgatgttccaaaatccaggaactacagtctggaggctcca	agcaaca	1979
D	b	1081	acctgcagatggtctgatgttccaaaatccaggaactacagtctggaggctcca	agcaaca	1140
Q	y	1980	aggaagttagtattggtagtatctcggtgttccatcgctgttagatgatctga	aaatgagac	2039
D	b	1141	aggaagttagtattggtagtatctcggtgttccatcgctgttagatgatctga	aaatgagac	1200
Q	y	2040	cgtctccattttgatgctctgggtttcgtccagatgattcacatgttccaatc	acggaataatcc	2099
D	b	1201	cgtctccattttgatgctctgggtttcgtccagatgattcacatgttccaatc	acggaataatcc	1360
Q	y	2100	tgaattctcaagctgcaccaacagagagctcgacacaaagctagacagcga	aaagcgcgtgg	2159
D	b	1261	tgaattctcaagctgcaccaacagagagctcgacacaaagctagacagcga	aaagcgcgtgg	1320
Q	y	2160	agatgacagtgcgtctgcagcgtggcgagatgctcagaaaactttagaagc	gctctagg	2219
D	b	1321	agatgacagtgcgtctgcagcgtggcgagatgctcagaaaactttagaagc	gctctagg	1380
Q	y	2220	taaaagctgggcaacaacagggcatactcaatgctttgggacagatcgctct	gcgtgcgtg	2279
D	b	1381	taaaagctgggcaacaacagggcatactcaatgctttaggacagatcgctct	gcgtgcgtg	1440
Q	y	2280	tgtagacgcaggaag-tccctccgcgtgcagcaagttctctatgg		2319
D	b	1441	tgtagacgcaggaagttccctccgcgtgcagcaagttctctatgg		1481

## RESULT 4

AAD02065  
ID AAD02065 standard; DNA; 1456 BP.

AA  
AC  
AAD02065:XX  
DT 26-MAR-2001 (first entry)

DE 3'-truncated Chlamydia pneumoniae 76 kDa protein encoding DNA.

76 kDa protein; bactericidal; diagnosis; prevention;  
 pneumonia; upper respiratory tract disease; sinusitis;  
 acute respiratory disease; cough; sore throat; hoarseness; fever;  
 vaccine; immunisation; treatment; truncation mutants.

OS Chlamydia pneumoniae.

OS	synthetic.	Location/Qualifiers
XX		
FH	Key	

FT	101..1456	/**tag= a	/product= "3'-truncated Chlamydia pneumoniae
FT			76kDa protein"
FT			/note= "The coding region does not include s
FT			/partial

AA  
PN  
WO200066739-A2

XX PD 09-NOV-2000.

XX  
PF 03-MAY-2000; 2000WO-CA00511.

XX  
PR 03-MAY-1999: 99US-0132270.

PR 30-JUN-1999; 99US-0141276.



601	agttaaagtgtggcgc	aaattacagaattagcga	aatatgttctgggataaac	caagcagat	660
1500	tcttgaccttttaggt	aaactgacttctctcag	actcttccacagactgt	ctcttctccaatc	1559
651	tcttgaccttttaggt	aaactgacttctctcag	actcttccacagactgt	ctcttctccaatc	720
1560	tgtagcaacaataac	aaagcagctgagctctt	ctaaagagatgc	caagataaacccagtagt	1619
721	tgtagcaacaataac	aaagcagctgagctctt	ctaaagagatgc	caagataaacccagtagt	780
1620	cccagggaacgcctc	gaattgctcaattttagt	tgtatcagacagatgc	tacacagcagac	1679
781	cccagggaacgcctc	gaattgctcaattttagt	tgtatcagacagatgc	tacacagcagac	840
1680	acagatagagaaga	tggaaatgcgattgggg	atgcataattttgc	aggacagaagcgtag	1739
841	acagatagagaaga	tggaaatgcgattgggg	atgcataattttgc	aggacagaagcgtag	900
1740	tggagctgtagaaa	tgtctaaatcaataac	agataaagcaacatag	atctcgactaaagc	1799
901	tggagctgtagaaa	tgtctaaatcaataac	agataaagcaacatag	atctcgactaaagc	960
1800	agcaatcgctactc	tctaagcacacaaa	tagctgaagctcaga	aaaagtcccgactctcc	1859
961	agcaatcgctactc	tctaagcacacaaa	tagctgaagctcaga	aaaagtcccgactctcc	1020
1860	aattcttcagaagc	ggacacaaatgtgt	taaacacagctgaga	aaagatcttaaaaatatcaa	1919
1021	aattcttcagaagc	ggacacaaatgtgt	taaacacagctgaga	aaagatcttaaaaatatcaa	1080
1920	acctcagatggttt	ctgatgttccaaatcc	aaagaaactcacagt	tggaggtctccaagcaaca	1979
1081	acctcagatggttt	ctgatgttccaaatcc	aaagaaactcacagt	tggaggtctccaagcaaca	1140
1980	aggaagtgtattgt	atgattcgtttccat	cgctgttagatgagt	ctgaaatcgagac	2039
1141	aggaagtgtattgt	atgattcgtttccat	cgctgttagatgagt	ctgaaatcgagac	1200
2040	cgtctccattttgat	gtctcgtgggtttc	gtccagatgatcca	atgttcaatacaggaaatcc	2099
1201	cgtctccattttgat	gtctcgtgggtttc	gtccagatgatcca	atgttcaatacaggaaatcc	1260
2100	tgatctcaagctgcc	caacaggagctcgc	gacagaagctagag	cagcagaaacccctgg	2159
1261	tgatctcaagctgcc	caacaggagctcgc	gacagaagctagag	cagcagaaacccctgg	1320
2160	agatgacagtgct	gtcgagcgctggc	agatgctcaga	aaagcttttagaagcgctctagg	2219
1321	agatgacagtgct	gtcgagcgctggc	agatgctcaga	aaagcttttagaagcgctctagg	1380
2220	taaagctggcaca	acaggggcactact	caatgctttggg	acagatcgctcttctgctgt	2279
1381	taaagctggcaca	acaggggcactact	caatgcttttag	acagatcgctcttctgctgt	1440
2280	tgtgagcgcagagt	2294			
1441	tgtgagcgcagagt	1455			

RESULT 5  
AAD2064  
ID AAD2064 standard; DNA; 1852 BP.  
XX  
XX  
AAD2064;  
AC  
XX  
XX  
26-MAR-2001 (first entry)  
XX  
XX  
XX  
DE 5'-truncated Chlamydia pneumoniae 76 kDa protein encoding DNA.  
XX  
KW 76 kDa protein; bactericidal; diagnosis; prevention;  
KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;  
KW acute respiratory disease; cough; sore throat; hoarseness; fever;  
KW vaccine; immunisation; treatment; truncation mutant; ds







```

FT CDS 101..1369
XX /*tag= a
XX WO200024901-A1.
XX 04-MAY-2000.
XX 28-OCT-1999; 99WO-GB03565.
XX 28-OCT-1998; 98US-0106037.
XX 20-SEP-1999; 99US-0154658.
XX 26-OCT-1999; 99US-0427501.
XX (CONN-) CONNAUGHT LAB LTD.
XX Murdin AD, Oomen RP, Dunn PL;
XX WPI; 2000-350742/30.
XX P-PSDB; AAY92716.
XX Isolated polynucleotide encoding a Chlamydia polypeptide useful to
XX treat, diagnose and prevent disease caused by Chlamydia infection
XX Claim 1; Fig 1A-C; 88pp; English.
XX This sequence encodes lorf2 protein of a strain of Chlamydia pneumoniae.
XX Comparison of this sequence as to the recently published genome sequence
XX of C. pneumoniae reveals that the sequence actually contains at least two
XX open reading frames, a first one in the 5' portion and a second one in
XX the 3' portion of the sequence. Despite the presence of the stop codon at
XX the end of this sequence, C. pneumoniae does make a 76 kDa product. It
XX appears possible that C. pneumoniae is able to read through this stop
XX codon and produce a full-length product terminated by the stop codon at
XX the end of the second open reading frame. There is at least one in-frame
XX ATG upstream of the start codon. This suggests that the first open
XX reading frame may form part of one or more larger open reading frames.
XX The lorf2 protein or DNA can be used as a vaccine for humans to treat or
XX prevent disease caused by Chlamydia infection. The sequences or an
XX antibody to lorf2 can be used to diagnose a Chlamydial infection.
XX Sequence 1550 BP; 472 A; 268 C; 317 G; 493 T; 0 other;

Query Match 32.7%; Score 832.8; DB 21; Length 1550;
Best Local Similarity 99.8%; Pred No. 1.2e-186;
Matches: 834; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 174 catgacaaaaaacattatgcttggtgtgtagaaggattctcaatgcttgcctaaaca 233
DB 715 catgacaaaaaacattatgcttggtgtgtagaaggattctcaatgcttgcctaaaca 774
QY 234 gtttttggaaatgtagtggctgcagctggaacacattcgtctccagaaacctccac 293
DB 775 gtttttggaaatgtagtggctgcagctggaacacattcgtctccagaaacctccac 834
QY 294 tacagaaaaagctgtcacaaacgctatgaatacacaatactgtgttgagctggtcgt 353
DB 835 tacagaaaaagctgtcacaaacgctatgaatacacaatactgtgttgagctggtcgt 894
QY 354 cggaaagcattgctcaggttcccttgatcaatggacagaaaaagcctctatctttatgg 413
DB 895 cggaaagcattgctcaggttcccttgatcaatggacagaaaaagcctctatctttatgg 954
QY 414 agctttcttaagaacccctttgcaaaagcgtacgaagctcgtttaaatggaaagaaaa 473
DB 955 agctttcttaagaacccctttgcaaaagcgtacgaagctcgtttaaatggaaagaaaa 1014
QY 474 cctagcttggtttattggaggaaactttaggggactcagaaagctggagactggtctgc 533
DB 1015 cctagcttggtttattggaggaaactttaggggactcagaaagctggagactggtctgc 1074
QY 534 cacagtcagttatgtagtgcgaagccttgcagttccagtaaaatagatgtttcaggat 593
DB 1075 cacagtcagttatgtagtgcgaagccttgcagttccagtaaaatagatgtttcaggat 1134
QY 594 tggcctggtaattattataaagtttggctgcgcccagcaattgctgctaactatgctcc 653
DB 1135 tggcctggtaattattataaagtttggctgcgcccagcaattgctgctaactatgctcc 1194
QY 654 taaagaggctaaatagttttacaaattataaaggattttccgctctatatatgtatgcat 713
DB 1195 taaagaggctaaatagttttacaaattataaaggattttccgctctatatatgtatgcat 1254
QY 714 cacagattctctatcattcagagcttatggggttactccaaacacagcaacgataaact 773
DB 1255 cacagattctctatcattcagagcttatggggttactccaaacacagcaacgataaact 1314
QY 774 cggcagtgattttactttccgaaagttagcttaggtataatttcagcgttttaagcaaa 833
DB 1315 cggcagtgattttactttccgaaagttagcttaggtataatttcagcgttttaagcaaa 1374
QY 834 atttaataaaatcttttaaaacaggtcgcatttaattatttagtgagagcgtttttttta 893
DB 1375 atttaataaaatcttttaaaacaggtcgcatttaattatttagtgagagcgtttttttta 1434
QY 894 tttttataataaaactaaagattttttatttttttttttttttttttttttttttttttt 953
DB 1435 tttttataataaaactaaagattttttatttttttttttttttttttttttttttttttt 1494
QY 954 tggctcaggtcctatagacaaacagcaacgacacacctccgcagatctttctgctc 1009
DB 1495 tggctcaggtcctatagacaaacagcaacgacacacctccgcagatctttctgctc 1550

RESULT 7
AA27110
ID AA27110 standard; cDNA; 1909 BP.
AC AA27110;
XX 04-AUG-2000 (first entry)
XX Rhesus monkey melanocortin-3-receptor cDNA.
XX Rhesus monkey; melanocortin-3 receptor protein; MC-3R;
XX G-protein coupled receptor; GPCR; obesity; ss.
XX Macaca mulatta.
XX Key Location/Qualifiers
XX CDS 148..1119
XX /*tag= a
XX /product= MC-3R protein
XX WO200027862-A1.
XX 18-MAY-2000.
XX 05-NOV-1999; 99WO-US25747.
XX 09-NOV-1998; 98US-0107725.
XX (MERI ) MERCK & CO INC.
XX Fong TM, Van Der Ploeg LHT, Huang RC;
XX WPI; 2000-387404/33.
XX P-PSDB; AAY94427.
XX New DNA encoding Rhesus monkey melanocortin 3 receptor protein,
XX recombinant vectors and host cells, useful in methods for identifying
XX selective agonists and antagonists -
XX Claim 1; Fig 1; 54pp; English.
XX A random primed probe from human melanocortin-3-receptor (MC-3R) was

```



CC used to screen a rhesus monkey genomic DNA library. One positive  
 CC lambda phage clone was identified. From this the rhesus  
 CC monkey MC-3R sequence was identified. Melanocortin receptors belong  
 CC to the rhodopsin sub-family of G-protein coupled receptors (GPCR).  
 CC The MC-3R protein and nucleotide sequence are useful for preparation  
 CC of recombinant host cells and antibodies. This is useful for  
 CC identifying ligand binding, activators and modulators, agonists and  
 CC antagonists of MC-3R. Antagonists of MC-3R may have use as novel  
 CC anti-obesity agents. The present sequence is rhesus monkey  
 CC melanocortin-3-receptor cDNA.

XX Sequence 1909 BP; 388 A; 581 C; 456 G; 484 T; 0 other;

Query Match 7.0%; Score 179; DB 21; Length 1909;  
 Best Local Similarity 99.5%; Pred. No. 1.6e-32;  
 Matches 190; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2347 gaacaaaactcatctcagaagagatctgaatagcgcgcgtgcacatcatcatcat 2406  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1720 gaacaaaactcatctcagaagagatctgaatagcgcgcgtgcacatcatcatcat 1779

QY 2407 cattaggttaaacggtctccagcttaagttaaacggtgcacgtcgcactgtgcct 2466  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1780 cattaggttaaacggtctccagcttaagttaaacggtgcacgtcgcactgtgcct 1839

QY 2467 tctagttgcagccatctgtttgtccctccctccctccctccctccctccctggaagt 2526  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1840 tctagttgcagccatctgtttgtccctccctccctccctccctccctccctggaagt 1898

QY 2527 gcaactccac 2537  
 Db |||||||||||  
 Db 1899 gcaactccac 1909

RESULT 8  
 AAX08423  
 ID AAX08423 standard; cDNA; 4912 BP.  
 AC AAX08423;  
 DT 28-JUN-1999 (first entry)  
 DE Recombinant vector VR-1012 comprising RSV G gene.  
 XX  
 KW G protein; respiratory syncytial virus; RSV; recombinant vector;  
 KW vaccine; immune response; immunogenicity; tPA; antibody;  
 KW tissue plasminogen activator; ss.  
 XX  
 OS Synthetic.  
 PN W09904010-A1.  
 XX  
 PD 28-JAN-1999.  
 XX  
 PF 16-JUL-1998; 98WO-CA00697.  
 XX  
 PR 18-JUL-1997; 97US-0896442.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Klein MH, Li X, Sambhara S;  
 XX  
 DR WPI; 1999-132254/11.  
 XX  
 PT Immunogenic composition for generating antibodies against  
 PT respiratory syncytial virus - comprises non-replicating vector  
 PT containing the protein G sequence, useful in protective vaccines and  
 PT to raise antibodies for diagnosis  
 XX  
 PS Example 1; Figure 6; 67pp; English.  
 XX  
 CC The respiratory syncytial virus (RSV) G protein can be used in

CC vaccines by inserting the G protein gene into a non-replicating  
 CC vector. The G protein is placed under the control of alternative  
 CC signal and expression sequences, for example the chimeric G protein  
 CC produced may also comprise the signal peptide of tissue plasminogen  
 CC activator (tPA). The recombinant vector may also comprise sequences  
 CC upstream of the G protein gene which enhance the G proteins  
 CC immunoprotective ability. The resulting immunogenic composition will  
 CC generate antibodies directed against the RSV G protein when  
 CC administered to a host organism. The composition is useful as a  
 CC vaccine to immunise against RSV-associated disease, particularly  
 CC resulting in a balanced Th1/Th2 immune response and for raising Ab,  
 CC by usual immunisation and cell fusion methods. VR-1012 is a  
 CC recombinant vector comprising the immediate-early cytomegalovirus  
 CC (CMV) promoter and Intron A sequences of human cytomegalovirus and  
 CC as a basis for the creating immunogenic vectors expressing the RSV  
 CC G protein. Two resulting vectors designated pXU5 and pXU6 have  
 CC been deposited with the American Type Culture Collection (ATCC  
 CC 209143 and ATCC 209144 respectively).

XX Sequence 4912 BP; 1241 A; 1235 C; 1188 G; 1248 T; 0 other;

Query Match 6.7%; Score 171; DB 20; Length 4912;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-30;  
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgcggtctgttaacggtgagggcagtgtagtgcagcagcagcagcagcagcagcagc 60  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1742 ttgcggtctgttaacggtgagggcagtgtagtgcagcagcagcagcagcagcagcagc 1801

QY 61 ggcgcacacagacataatagctgacagactaacagactgtctcttccatgggtctttct 120  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1802 ggcgcacacagacataatagctgacagactaacagactgtctcttccatgggtctttct 1861

QY 121 gcagtcacgctgcgcacagcgtgtgatcagatcgcgcgcgcgcgcgcgcgcgcgcgc 171  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1862 gcagtcacgctgcgcacagcgtgtgatcagatcgcgcgcgcgcgcgcgcgcgcgcgc 1912

RESULT 9  
 AAA71402  
 ID AAA71402 standard; DNA; 4912 BP.  
 XX  
 AC AAA71402;  
 DT 01-DEC-2000 (first entry)  
 XX  
 DE plasmid VR-1012 DNA.  
 XX  
 KW F protein; immunization; vaccine; infection; seronegative;  
 KW anti-F antibody; viral antigen; cytotoxic T-cell induction; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN US06083925-A.  
 XX  
 PD 04-JUL-2000.  
 XX  
 PF 05-MAR-1999; 99US-0262927.  
 XX  
 PR 07-JUN-1995; 95US-0476397.  
 XX  
 PR 07-JUN-1996; 96US-0659939.  
 XX  
 PR 18-JUL-1997; 97US-0896500.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Klein ME, Sambhara S, Ewasyshyn ME, Li X;  
 XX  
 DR WPI; 2000-464371/40.  
 XX  
 PT Immunizing a host against diseases caused by respiratory syncytial  
 PT virus (RSV) infection comprises administering a non-replicating vector



PT having a nucleotide sequence coding for RSV F protein and a promoter of  
PT the sequence

Example 7; Fig 11; 61pp; English.

XX This invention describes a novel method for immunizing a host against  
CC diseases caused by infection with RSV which comprises administering a  
CC plasmid vector comprising a nucleotide sequence encoding an RSV F protein  
CC fragment, and a promoter sequence operatively coupled to the nucleotide  
CC sequence. The RSV F protein generates antibodies and/or cytotoxic  
CC T-lymphocytes (CTLs) that specifically react with the RSV F protein. The  
CC method is useful for administering RSV vaccines to seronegative infants.  
CC The plasmid vector is useful for immunizing against RSV infection, as  
CC well as in diagnostic applications. The RSV F genes and vectors are also  
CC useful as immunogens for the generation of anti-F antibodies for use in  
CC immunoassays (e.g. enzyme linked immunosorbent assay (ELISA),  
CC radioimmunoassay (RIA) or other non-enzyme linked antibody binding  
CC assays). Prior methods of administering RSV vaccines have proven to be  
CC poorly immunogenic with regard to the induction of neutralizing  
CC antibodies in seronegative humans or chimpanzees. These methods also  
CC cause disease enhancement or immunopotential, and have the risk of  
CC residual virulence and genetic instability. In the present method, the  
CC DNA encoding a viral antigen can be introduced in the presence of  
CC antibody to the virus itself, without loss of potency due to  
CC neutralization of virus by the antibodies. The antigen expressed in vivo  
CC exhibits a native conformation. Therefore, it induces an antibody  
CC response similar to that induced by the antigen present in the wild-type  
CC virus infection. The expression of proteins from injected plasmid DNA can  
CC be detected in vivo for a considerably longer period than in  
CC virus-infected cells. This is advantageous since it prolongs cytotoxic  
CC T-cell induction and enhances antibody responses. The in vivo expression  
CC of antigen may also provide protection without the need for an extrinsic  
CC adjuvant. This sequence the encoding DNA sequence of the plasmid VR-1012  
CC which is used in the method of the invention.

XX Sequence 4912 BP; 1241 A; 1236 C; 1188 G; 1247 T; 0 other;

Query Match 6.7%; Score 171; DB 21; Length 4912;  
Best Local Similarity 100.0%; Pred. No. 1.8e-30;  
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagtagctgttgcgcgcgc 60  
Db 1742 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagtagctgttgcgcgcgc 1801  
QY 61 gcgccaccagacataatagctgacagactaacagactgttccattccatgggtttttct 120  
Db 1802 gcgccaccagacataatagctgacagactaacagactgttccattccatgggtttttct 1861  
QY 121 gcagtcacgctgctgacacgctgtgatcagatcgccgctctagacc 171  
Db 1862 gcagtcacgctgctgacacgctgtgatcagatcgccgctctagacc 1912

RESULT 10  
AAC86258  
ID AAC86258 standard; DNA; 5108 BP.  
XX AAC86258;  
XX AAC86258;  
XX 28-FEB-2001 (first entry)  
DE Plasmid GHRH1-29WTCMV.  
XX GHRH; growth hormone-releasing hormone; enzyme degradation; ds.  
XX Synthetic.  
XX EP1052286-A2.  
XX 15-NOV-2000.

PF 12-APR-2000; 2000EP-0302790.  
XX 12-APR-1999; 99US-0128830.  
XX (PFIZ ) PFIZER PROD INC.  
XX Morsey MA, Sheppard MG;  
PI WPI; 2001-026585/04.  
XX New polypeptide variants of growth hormone releasing hormone with  
XX enhanced resistance to enzymatic degradation, useful for treating  
XX growth hormone deficiency related disorders or to improve growth and  
XX performance

XX Examples; Page 39-41; 67pp; English.

XX The present invention relates to growth hormone-releasing hormone  
CC (GHRH) variants having resistance to enzymatic degradation.  
CC The variant GHRH polypeptides can be administered to animals to  
CC treat growth hormone deficiency related disorders, or to improve  
CC growth and/or performance. The variants can be included in  
CC pharmaceutical compositions to promote expression and elevation of  
CC growth hormone. The variants can be produced recombinantly at much  
CC higher levels than prior art variants modified using traditional  
CC chemical methods. They have enhanced resistance to enzymatic  
CC degradation, therefore have increased length of activity.  
XX Sequence 5108 BP; 1279 A; 1312 C; 1231 G; 1286 T; 0 other;

Query Match 6.7%; Score 171; DB 22; Length 5108;  
Best Local Similarity 100.0%; Pred. No. 1.8e-30;  
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagtagctgttgcgcgcgc 60  
Db 4723 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagtagctgttgcgcgcgc 4782  
QY 61 gcgccaccagacataatagctgacagactaacagactgttccattccatgggtttttct 120  
Db 4783 gcgccaccagacataatagctgacagactaacagactgttccattccatgggtttttct 4842  
QY 121 gcagtcacgctgctgacacgctgtgatcagatcgccgctctagacc 171  
Db 4843 gcagtcacgctgctgacacgctgtgatcagatcgccgctctagacc 4893

RESULT 11  
AAC86259  
ID AAC86259 standard; DNA; 5108 BP.  
XX AAC86259;  
XX AAC86259;  
XX 28-FEB-2001 (first entry)  
XX Plasmid GHRH1-29WTCMV.  
XX GHRH; growth hormone-releasing hormone; enzyme degradation; ds.  
XX Synthetic.  
XX EP1052286-A2.  
XX 15-NOV-2000.  
XX 12-APR-2000; 2000EP-0302790.  
XX 12-APR-1999; 99US-0128830.  
XX (PFIZ ) PFIZER PROD INC.  
XX Morsey MA, Sheppard MG;  
PI



XX WPI; 2001-026585/04.  
 XX New polypeptide variants of growth hormone releasing hormone with  
 PT enhanced resistance to enzymatic degradation, useful for treating  
 PT growth hormone deficiency related disorders or to improve growth and  
 PT performance -  
 XX  
 PS Examples; Page 41-43; 67pp; English.  
 XX The present invention relates to growth hormone-releasing hormone  
 CC (GHRH) variants having enhanced resistance to enzymatic degradation.  
 CC The variant GHRH polypeptides can be administered to animals to  
 CC treat growth hormone deficiency related disorders, or to improve  
 CC growth and/or performance. The variants can be included in  
 CC pharmaceutical compositions to promote expression and elevation of  
 CC growth hormone. The variants can be produced recombinantly at much  
 CC higher levels than prior art variants modified using traditional  
 CC chemical methods. They have enhanced resistance to enzymatic  
 CC degradation, therefore have increased length of activity.  
 XX  
 SQ Sequence 5108 BP; 1279 A; 1313 C; 1230 G; 1286 T; 0 other;  
 Query Match 6.7%; Score 171; DB 22; Length 5108;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-30;  
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactcgttctccatgggtctttct 120  
 DB 4723 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactcgttctccatgggtctttct 4842  
 QY 61 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 120  
 DB 4783 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 4842  
 QY 121 gcagtcacgctgctgacacgctgtgatcagatatcgcgccgctctagacc 171  
 DB 4843 gcagtcacgctgctgacacgctgtgatcagatatcgcgccgctctagacc 4893  
 RESULT 12  
 AAC86262  
 ID AAC86262 standard; DNA; 5111 BP.  
 AC AAC86262;  
 XX  
 XX 28-FEB-2001 (first entry)  
 DE CMV IE promoter enhancer elements.  
 XX  
 KW GHRH; growth hormone-releasing hormone; enzyme degradation; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN EP1052286-A2.  
 XX  
 PD 15-NOV-2000.  
 XX  
 PF 12-APR-2000; 2000EP-0302790.  
 XX  
 PR 12-APR-1999; 99US-0128830.  
 XX  
 PA (PFIZ ) PFIZER PROD INC.  
 XX  
 PI Morsey MA, Sheppard MG;  
 XX  
 DR WPI; 2001-026585/04.  
 XX  
 PT New polypeptide variants of growth hormone releasing hormone with  
 PT enhanced resistance to enzymatic degradation, useful for treating  
 PT growth hormone deficiency related disorders or to improve growth and  
 PT performance -

XX  
 PS Examples; Page 47-49; 67pp; English.  
 XX The present invention relates to growth hormone-releasing hormone  
 CC (GHRH) variants having enhanced resistance to enzymatic degradation.  
 CC The variant GHRH polypeptides can be administered to animals to  
 CC treat growth hormone deficiency related disorders, or to improve  
 CC growth and/or performance. The variants can be included in  
 CC pharmaceutical compositions to promote expression and elevation of  
 CC growth hormone. The variants can be produced recombinantly at much  
 CC higher levels than prior art variants modified using traditional  
 CC chemical methods. They have enhanced resistance to enzymatic  
 CC degradation, therefore have increased length of activity.  
 XX  
 SQ Sequence 5111 BP; 1280 A; 1314 C; 1230 G; 1287 T; 0 other;  
 Query Match 6.7%; Score 171; DB 22; Length 5111;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-30;  
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactcgttctccatgggtctttct 120  
 DB 4723 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactcgttctccatgggtctttct 4782  
 QY 61 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 120  
 DB 4783 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 4842  
 QY 121 gcagtcacgctgctgacacgctgtgatcagatatcgcgccgctctagacc 171  
 DB 4843 gcagtcacgctgctgacacgctgtgatcagatatcgcgccgctctagacc 4893  
 RESULT 13  
 AAC86254  
 ID AAC86254 standard; DNA; 5185 BP.  
 AC AAC86254;  
 XX  
 XX 28-FEB-2001 (first entry)  
 DE pGHRH-4 construct.  
 XX  
 KW GHRH; growth hormone-releasing hormone; enzyme degradation; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN EP1052286-A2.  
 XX  
 PD 15-NOV-2000.  
 XX  
 PF 12-APR-2000; 2000EP-0302790.  
 XX  
 PR 12-APR-1999; 99US-0128830.  
 XX  
 PA (PFIZ ) PFIZER PROD INC.  
 XX  
 PI Morsey MA, Sheppard MG;  
 XX  
 DR WPI; 2001-026585/04.  
 XX  
 PT New polypeptide variants of growth hormone releasing hormone with  
 PT enhanced resistance to enzymatic degradation, useful for treating  
 PT growth hormone deficiency related disorders or to improve growth and  
 PT performance -  
 XX  
 PS Examples; Page 32-34; 67pp; English.  
 XX The present invention relates to growth hormone-releasing hormone  
 CC (GHRH) variants having enhanced resistance to enzymatic degradation.  
 CC The variant GHRH polypeptides can be administered to animals to  
 CC treat growth hormone deficiency related disorders, or to improve



Query Match 6.7%; Score 171; DB 22; Length 5188;  
Best Local Similarity 100.0%; Pred. No. 1.8e-30;  
Matches 171: Conservative 0; Mismatches 0; Indels 0

Qy 1 ttgcggtgctgttaacagtggaaggcagtgtagctgagcagtaactcgttgctgcgcgcg 60  
:  
Dd 4723 ttgcggtgctgttaacagtggaaggcagtgtagctgagcagtaactcgttgctgcgcgcg 4782

[illegible]

DW 4763 gcgcacacacagacataatagctgcacagatccacacagatcgtccccccccccatggggcccccccc 484  
 QY 121 gcagtcaccgtcgtcgacacgctgtgatcagatatcgcgcgcgcctctagacc 171

**Q7** 121 gcagtcacccgcgctcgatcgcgcgtgatcgcgcgtctctagacc 171  
|||||  
**Db** 4843 gcagtaccgctggctgcacgcgtgatcagatatcgcgcgcgtctctagacc 4893

RESULT 15

RESULT 15  
AAC86267  
ID AAC86267 standard: DNA: 5254 BP.

AAC00207  
 ID AAC86267 standard; DNA; 5254 BP.  
 XX  
 AC AAC86267;

AA	
AC	AAC86267;
XX	
DT	28-FEB-2001 (first entry)

DT 28-FEB-2001 (FIRST ENTRY)  
XX

Plasmid pGHRH1-44WTGHeep.

XX  
KW  
MY  
GHK; growth hormone-rele

XX  
XX  
содержит.

PN EPT052200-AZ.  
XX

FD  
T3-NOV-2000.  
XX

XX  
12-APR-2000XX  
PR 12-APR-1959;XX  
FA (FEL) FELTPT MORSEY MA, SHEPPARD MG;  
XX

DR WP1; 2001-020383/04.  
XX

PT New polypeptide variants enhanced resistance to e

PT growth hormone deficiency -  
PT performance -

Examples: Page 56-58

The present invention relates to growth hormone-releasing hormone (GHRH) variants having enhanced resistance to enzymatic degradation. The variant GHRH polypeptides can be administered to animals to treat growth hormone deficiency related disorders, or to improve

growth and/or performance. The value of pharmaceutical compositions to promote

CC growth hormone, the variants can be produced recombinantly at much  
CC higher levels than prior art variants modified using traditional  
CC chemical methods. They have enhanced resistance to enzymatic  
CC degradation, therefore have increased length of activity.  
XX  
SQ Sequence 5254 BP; 1318 A; 1334 C; 1285 G; 1317 T; 0 other;

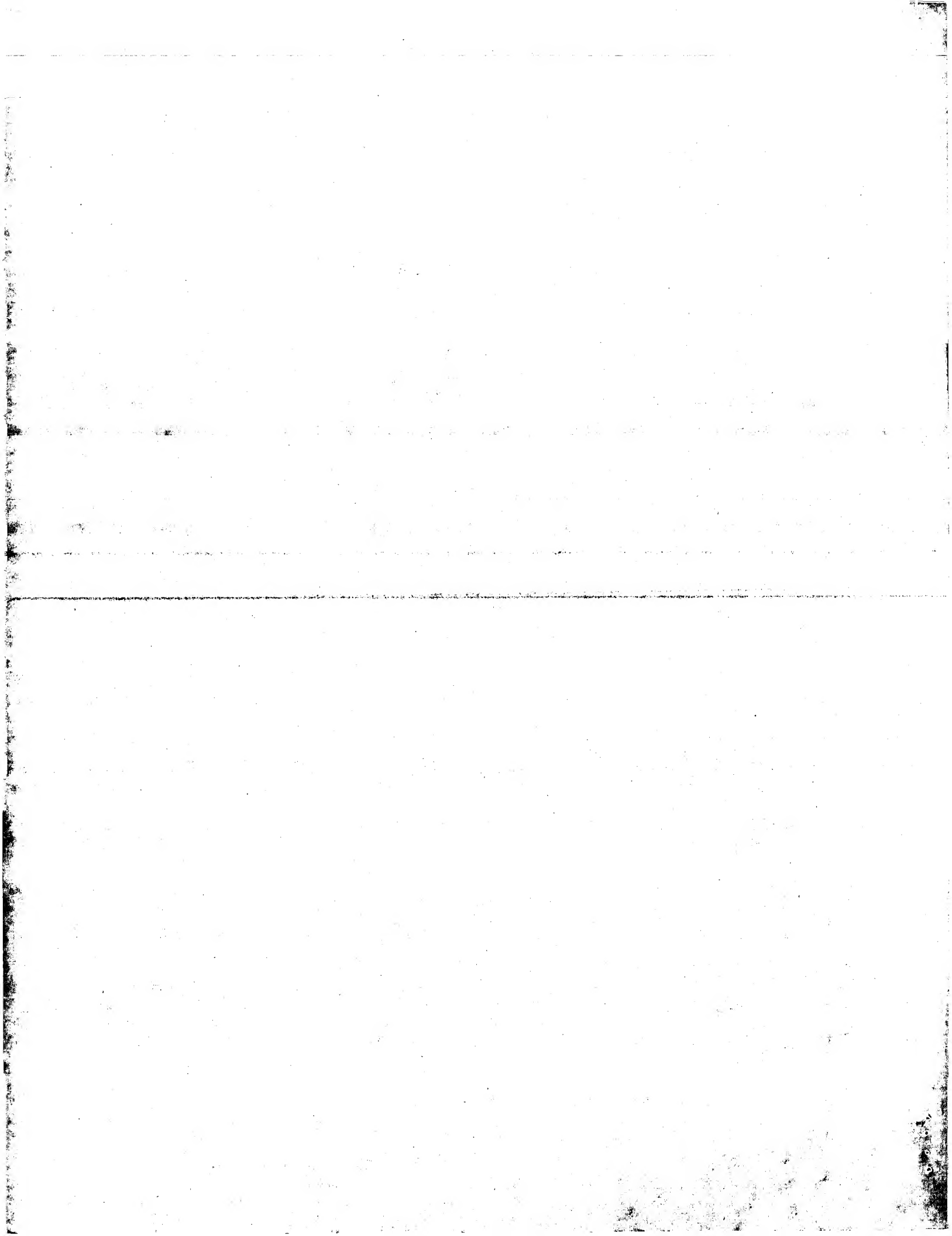
```
Query Match          6.7%; Score 171; DB 22; Length 5254;
Best Local Similarity 100.0%; Pred. No. 1.8e-30;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



QY 1 ttgcggtgctgttaacggtggaggcagtgtagtctgagcagtagtctgctgcgcgc 60  
 Db 4723 ttgcggtgctgttaacggtggaggcagtgtagtctgagcagtagtctgctgcgcgc 4782  
 QY 61 gcgccaccagacataatagctgacagactaacagactgttcccttccatgggtctttct 120  
 Db 4783 gcgccaccagacataatagctgacagactaacagactgttcccttccatgggtctttct 4842  
 QY 121 gcagtcaccgctgctgacacgctgtgtagtcagatatacgcgccgctctagacc 171  
 Db 4843 gcagtcaccgctgctgacacgctgtgtagtcagatatacgcgccgctctagacc 4893

Search completed: February 7, 2002, 19:46:11  
 Job time: 14285 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:48:06 ; Search time 243.49 Seconds  
(without alignments)  
2367.188 Million cell updates/sec

Title: US-09-391-606-1  
Perfect score: 2545  
Sequence: 1 ttgcgggtgcttaacgggtg.....tgccactccactgtccttt 2545

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149.2	5.9	4928	1	US-08-345-913-1
2	149.2	5.9	4928	3	US-08-818-562-1
3	146.8	5.8	150	4	US-08-659-473-8
4	142	5.6	3125	2	US-08-037-816A-13
5	142	5.6	3125	2	US-08-530-146-13
6	139.8	5.5	4328	4	US-09-132-808-1
7	139.8	5.5	4328	4	US-08-910-647-2
8	139.8	5.5	4818	4	US-08-910-647-4
9	139.8	5.5	5107	4	US-08-910-647-3
10	139.8	5.5	9600	4	US-08-910-647-1
11	138.2	5.4	5676	2	US-08-663-998-3
12	138.2	5.4	5682	2	US-08-663-998-4
13	138.2	5.4	5900	2	US-08-663-998-1
14	138.2	5.4	5952	2	US-08-663-998-2
15	138	5.4	2057	1	US-08-450-945-57
16	138	5.4	2057	1	US-08-976-161-57
17	137.8	5.4	13254	1	US-08-276-852-156
18	137.8	5.4	13254	1	US-08-276-852-170
19	137.8	5.4	13254	1	US-08-899-575-156
20	137.8	5.4	13254	1	US-08-899-575-170
21	137.8	5.4	13254	1	US-08-899-575-156
22	137.8	5.4	13254	1	US-08-899-575-170
23	137.8	5.4	13254	5	PCT-US95-08743-156
24	137.8	5.4	13254	5	PCT-US95-08743-170
25	132	5.2	4326	4	US-08-760-615-7
26	111.6	4.4	420	4	US-09-158-863C-64
27	109	4.3	456	1	US-08-450-945-70

Sequence 70, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 19, Appl  
Sequence 33, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 3, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 15, Appl  
Sequence 17, Appl  
Sequence 19, Appl  
Sequence 1, Appl

456 4 US-08-976-161-70  
6253 2 US-08-627-151A-5  
3853 3 US-08-801-092-5  
4026 3 US-08-801-092-19  
4249 3 US-08-801-092-33  
4283 1 US-08-343-401A-3  
4283 1 US-08-445-265A-1  
4283 3 US-08-990-442-1  
4965 2 US-08-564-313-1  
4965 5 PCT-US94-06069-1  
5653 1 US-08-073-836-3  
5653 1 US-08-235-277-1  
6238 3 US-08-646-538-5  
6238 4 US-09-503-222-5  
6253 3 US-08-893-327-15  
6280 3 US-08-893-327-17  
6280 3 US-08-893-327-19  
6367 1 US-08-470-299-1

## ALIGNMENTS

RESULT 1  
US-08-345-913-1  
; Sequence 1, Application US/08345913  
; Patent No. 5641665  
; GENERAL INFORMATION:  
; APPLICANT: Hobart, Peter  
; APPLICANT: Parker, Suzanne  
; APPLICANT: Margalith, Michal  
; APPLICANT: Khatibi, Shirin  
; TITLE OF INVENTION: PLASMIDS SUITABLE FOR IL-2 EXPRESSION  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobber, Martens, Olson and Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/345,913  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ways Vensko, Nancy  
; REGISTRATION NUMBER: 36,298  
; REFERENCE/DOCKET NUMBER: VICAL.043A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-235-8550  
; TELEFAX: 619-235-0176  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4928 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; FEATURE:







COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/037,816A  
FILING DATE: 26-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41190/JPW/AJM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3125 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1555..3115  
OTHER INFORMATION:  
US-08-037-816A-13

Query Match 5.6%; Score 142; DB 2; Length 3125;  
Best Local Similarity 91.0%; Pred. No. 1.2e-28;  
Matches 151; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ttgcggtcgtgttaacggtgagggcagtgtagtctgagcagctactcgttgcgcgcgc 60  
|||||  
DB 1414 TTGCGGTGCTGTTAAAGGTGGAGGCGAGTGTAGTCTGAGCAGTACTCGTTGTCGCCGC 1473  
QY 61 ggcgcacagacataatagctgacagactaacagactgttccttccatgggtctttct 120  
|||||  
DB 1474 GCGCCACCAGACATAATAGTCTGACAGACTTAACAGAGACTGTTCCTCCATGGGTCTTTCT 1533  
QY 121 cgagtcaccgtcgtcgtgacagctgtgatcagatcgcgcgcgcctct 166  
|||||  
DB 1534 GCAGTCACCGTCTTGTACACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1579

RESULT 5  
US-08-146-13  
Sequence 13, Application US/08530146  
Patent No. 5886163  
GENERAL INFORMATION:  
APPLICANT: Hasel, Karl W.  
TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED  
THEREOF, AND THERAPEUTIC AND PROPHYLACTIC USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/530,146  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/037,816  
FILING DATE: 26-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41190/JPW/AJM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOPUI  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3125 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1555..3115  
OTHER INFORMATION:  
US-08-530-146-13

Query Match 5.6%; Score 142; DB 2; Length 3125;  
Best Local Similarity 91.0%; Pred. No. 1.2e-28;  
Matches 151; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ttgcggtcgtgttaacggtgagggcagtgtagtctgagcagctactcgttgcgcgcgc 60  
|||||  
DB 1414 TTGCGGTGCTGTTAAAGGTGGAGGCGAGTGTAGTCTGAGCAGTACTCGTTGTCGCCGC 1473  
QY 61 ggcgcacagacataatagctgacagactaacagactgttccttccatgggtctttct 120  
|||||  
DB 1474 GCGCCACCAGACATAATAGTCTGACAGACTTAACAGAGACTGTTCCTCCATGGGTCTTTCT 1533  
QY 121 cgagtcaccgtcgtcgtgacagctgtgatcagatcgcgcgcgcctct 166  
|||||  
DB 1534 GCAGTCACCGTCTTGTACACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1579

RESULT 6  
US-09-132-803-1  
Sequence 1, Application US/09132808  
Patent No. 6197332  
GENERAL INFORMATION:  
APPLICANT: Ronald Zuckermann et al.

TITLE OF INVENTION: Lipid-Conjugated Polyamide Compounds and Related  
Compositions and Methods Thereof  
TITLE OF INVENTION: COMPOSITIONS AND METHODS THEREOF  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94608-2916  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/132,808  
FILING DATE:  
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Fujita, Sharon M.  
REGISTRATION NUMBER: 38,459  
REFERENCE/DOCKET NUMBER: 1387,002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 923-2706



TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4328 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-132-808-1

Query Match 5.5%; Score 139.8; DB 4; Length 4328;  
Best Local Similarity 95.4%; Pred. No. 5.6e-28;  
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagctgacagactgttccttccatgggtctttct 60  
Db 1511 TTGCGGTGCTGTTAAACGGTGGAGGCGAGTGTAGTCTGAGCAGTACTCGTTGCGCGCGC 1570  
QY 61 gcgccaccagacataatagctgacagactaacagagactgttccttccatgggtctttct 120  
Db 1571 GCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCCTTCCATGGGTCTTTCT 1630  
QY 121 gcagtcaccgtcgtagcagcgtgtgatcaga 151  
Db 1631 GCAGTCACCGCTCGTCCGACCTAAGAAATTCAGA 1661

RESULT 8  
US-08-910-647-4  
; Sequence 2, Application US/08910647  
; Patent No. 6251433  
; GENERAL INFORMATION:  
; APPLICANT: Zuckermann et al.  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Polynucleotide Delivery  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/08/910,647  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fujita, Sharon M.  
; REGISTRATION NUMBER: 38,459  
; REFERENCE/DOCKET NUMBER: 1218.002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-2706  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4818 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-910-647-4

Query Match 5.5%; Score 139.8; DB 4; Length 4818;  
Best Local Similarity 95.4%; Pred. No. 6e-28;  
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagctgacagactgttccttccatgggtctttct 60  
Db 1514 TTGCGGTGCTGTTAAACGGTGGAGGCGAGTGTAGTCTGAGCAGTACTCGTTGCGCGCGC 1573  
QY 61 gcgccaccagacataatagctgacagactaacagagactgttccttccatgggtctttct 120  
Db 1574 GCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCCTTCCATGGGTCTTTCT 1633  
QY 121 gcagtcaccgtcgtagcagcgtgtgatcaga 151  
Db 1634 GCAGTCACCGCTCGTCCGACCTAAGAAATTCAGA 1664

Query Match 5.5%; Score 139.8; DB 4; Length 4328;  
Best Local Similarity 95.4%; Pred. No. 5.6e-28;  
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;



## RESULT 9

US-08-910-647-3  
; Sequence 3, Application US/08910647  
; Patent No. 6251433  
; GENERAL INFORMATION:  
; APPLICANT: Zuckermann et al.  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Polynucleotide Delivery  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910.647  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fujita, Sharon M.  
; REGISTRATION NUMBER: 38,459  
; REFERENCE/DOCKET NUMBER: 1218.002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-2706  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5107 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-910-647-3

Query Match 5.5%; Score 139.8; DB 4; Length 5107;  
Best Local Similarity 95.4%; Pred. No. 6.le-28;  
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 1 ttgcggtctgttaacggtgagggcagtgtagctgagcagtagctgttcttccatgggtcttttct 120  
Db 1514 TTGCGGTGCTGTTAACGGTGGAGGCGAGTGTAGCTGACAGACTGTTCTTTCCATGGGTCTTTCT 1573  
Qy 61 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtcttttct 120  
Db 1574 GCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCTTTCCATGGGTCTTTCT 1633  
Qy 121 gcagtcacgctgctgacacgctgtgatcaga 151  
Db 1634 GCAGTCACCGTGGTGGACCTAAGAAATTCAGA 1664

## RESULT 10

US-08-910-647-1  
; Sequence 1, Application US/08910647  
; Patent No. 6251433  
; GENERAL INFORMATION:  
; APPLICANT: Zuckermann et al.  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Polynucleotide Delivery  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California

COUNTRY: U.S.A.  
ZIP: 94608-2916  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910.647  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Fujita, Sharon M.  
REGISTRATION NUMBER: 38,459  
REFERENCE/DOCKET NUMBER: 1218.002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 923-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9600 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-910-647-1  
Query Match 5.5%; Score 139.8; DB 4; Length 9600;  
Best Local Similarity 95.4%; Pred. No. 8.6e-28;  
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 1 ttgcggtctgttaacggtgagggcagtgtagctgagcagtagctgttcttccatgggtcttttct 60  
Db 6616 TTGCGGTGCTGTTAACGGTGGAGGCGAGTGTAGCTGACAGACTGTTCTTTCCATGGGTCTTTCT 6675  
Qy 61 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtcttttct 120  
Db 6676 GCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCTTTCCATGGGTCTTTCT 6735  
Qy 121 gcagtcacgctgctgacacgctgtgatcaga 151  
Db 6736 GCAGTCACCGTGGTGGACCTAAGAAATTCAGA 6766

## RESULT 11

US-08-663-998-3  
; Sequence 3, Application US/08663998  
; Patent No. 5846946  
; GENERAL INFORMATION:  
; APPLICANT: HUEBNER, Robert C.  
; APPLICANT: NORMAN, Jon A.  
; APPLICANT: LIANG, Xiaowu  
; APPLICANT: CARNER, Kristin R.  
; APPLICANT: BARBOUR, Alan G.  
; APPLICANT: LUKE, Catherine J.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/663.998



```

; FILING DATE: 06-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalski, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454312-2440
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5676 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-663-998-3

Query Match          5.4%; Score 138.2; DB 2; Length 5676;
Best Local Similarity 97.9%; Pred. No. 1.8e-27;
Matches 140; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagtagtctggtgctgctccgcgc 60
Db 1549 TTGCGGTGCTGTTAACGGTGGAGGCGAGTGTAGTCTGAGCAGTACTCTGTTGCTGCGCGC 1608

QY 61 ggcaccagacataatagctgacagactaacagactgttcttccatgggtctttct 120
Db 1609 GCGCCACCAGACATAATAGCTGACAGACTTAACAGACTGTTCCTTCCATGGGTCTTTTCT 1668

QY 121 gcagtcaccgtctgcacacgtg 143
Db 1669 GCAGTCACCGTCTGCGACAGAG 1691

```

```

RESULT 12
US-08-663-998-4
; Sequence 1, Application US/08663998
; Patent No. 5846946
; GENERAL INFORMATION:
; APPLICANT: HUEBNER, Robert C.
; APPLICANT: NORMAN, Jon A.
; APPLICANT: LIANG, Xiaowu
; APPLICANT: CARNER, Kristin R.
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: LUKE, Catherine J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,998
; FILING DATE: 06-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalski, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454312-2440
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 5682 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-663-998-4

Query Match          5.4%; Score 138.2; DB 2; Length 5682;
Best Local Similarity 97.9%; Pred. No. 1.8e-27;
Matches 140; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagtagtctggtgctgctccgcgc 60
Db 1549 TTGCGGTGCTGTTAACGGTGGAGGCGAGTGTAGTCTGAGCAGTACTCTGTTGCTGCGCGC 1608

QY 61 ggcaccagacataatagctgacagactaacagactgttcttccatgggtctttct 120
Db 1609 GCGCCACCAGACATAATAGCTGACAGACTTAACAGACTGTTCCTTCCATGGGTCTTTTCT 1668

QY 121 gcagtcaccgtctgcacacgtg 143
Db 1669 GCAGTCACCGTCTGCGACAGAG 1691

```

```

RESULT 13
US-08-663-998-1
; Sequence 1, Application US/08663998
; Patent No. 5846946
; GENERAL INFORMATION:
; APPLICANT: HUEBNER, Robert C.
; APPLICANT: NORMAN, Jon A.
; APPLICANT: LIANG, Xiaowu
; APPLICANT: CARNER, Kristin R.
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: LUKE, Catherine J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,998
; FILING DATE: 06-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalski, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454312-2440
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5900 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-663-998-1

Query Match          5.4%; Score 138.2; DB 2; Length 5900;
Best Local Similarity 97.9%; Pred. No. 1.8e-27;

```



Search completed: February 7, 2002, 15:53:24  
Job time: 318 sec

Search completed: February 7, 2003 1

Search completed: February 7, 2002, 15:53:24  
Job time: 318 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:48:06 ; Search time 7419.31 Seconds  
(without alignments)  
3686.058 Million cell updates/sec

Title: US-09-391-606-1

Perfect score: 2545

Sequence: 1 ttgcgggtgcttaacgggtg.....tgccactcccaactgtctcttt 2545

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rnd:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	3.4	149	10 AW669676	AW669676 113173 MA
2	87	3.4	162	10 AW669522	AW669522 112928 MA
3	87	3.4	162	10 AW785253	AW785253 116166 MA
4	87	3.4	162	10 AW785901	AW785901 117399 MA
5	87	3.4	243	10 AW669518	AW669518 112921 MA
6	87	3.4	503	11 BF074689	BF074689 22166 MA
7	85.4	3.4	328	10 AW669634	AW669634 113105 MA
8	84.4	3.3	374	10 AW669633	AW669633 113104 MA
9	71	2.8	98	11 BG365021	BG365021 100689 MA
10	61.2	2.4	1001	13 CNS007BE	AL066953 Drosophil
11	60	2.4	1101	13 CNS0039G	AL063921 Drosophil
12	59.4	2.3	996	13 CNS00FUH	AL071063 Drosophil

C 13	56	2.2	987	13 CNS014PQ	AL104456 Drosophil
C 14	52.8	2.1	976	13 CNS04E5M	AL286627 Tetraodon
C 15	52.4	2.1	240	10 AU071789	AU071789 Drosophil
C 16	52.2	2.1	928	13 CNS00DKY	AL071865 Drosophil
C 17	52.2	2.1	937	13 CNS006XP	AL066056 Drosophil
C 18	51.4	2.0	663	13 BH001815	AL001815 BMBAC01K0
C 19	51.2	2.0	1190	13 CNS020N7	AL206908 Tetraodon
C 20	51	2.0	1082	13 CNS012XZ	AL102161 Drosophil
C 21	51	2.0	1092	13 CNS020K7	AL175696 Tetraodon
C 22	51	2.0	1101	13 CNS00FYG	AL071206 Drosophil
C 23	50.8	2.0	1001	13 CNS0155H	AL105023 Drosophil
C 24	50.6	2.0	1225	13 CNS0161D	AL106171 Drosophil
C 25	50.4	2.0	842	13 AZ670257	AZ670257 ENTKS53TF
C 26	50.4	2.0	1101	13 CNS003B0	AL064078 Drosophil
C 27	50.2	2.0	822	13 CNS009CW	AL053618 Drosophil
C 28	50.2	2.0	1101	13 CNS00K55	AL077453 Drosophil
C 29	50.2	2.0	1225	13 CNS0161D	AL106171 Drosophil
C 30	50	2.0	842	13 CNS010QN	AL093305 Drosophil
C 31	45.8	2.0	942	13 CNS018GS	AL109318 Drosophil
C 32	45.8	2.0	1101	13 CNS000B8	AL063632 Drosophil
C 33	45.6	1.9	668	10 AL514901	AL514901 AL514901
C 34	45.4	1.9	777	13 CNS025WB	AL184612 Tetraodon
C 35	45.4	1.9	1083	13 CNS009YQ	AL054505 Drosophil
C 36	45.2	1.9	408	13 AQ954772	AQ954772 nbab0074A
C 37	45.2	1.9	1131	13 CNS034FO	AL227373 Tetraodon
C 38	45.2	1.9	1203	13 CNS015WU	AL106008 Drosophil
C 39	49	1.9	905	13 CNS00KHX	AL077798 Drosophil
C 40	48.8	1.9	1101	13 CNS00LVZ	AL079819 Drosophil
C 41	48.8	1.9	1101	13 CNS016DT	AL106619 Drosophil
C 42	48.6	1.9	838	13 CNS03272	AL267272 Tetraodon
C 43	48.6	1.9	988	13 CNS0161M	AL106180 Drosophil
C 44	48.4	1.9	773	13 AZ183506	AZ183506 SP_1001_A
C 45	48.4	1.9	945	13 CNS04D0K	AL285149 Tetraodon

# ALIGNMENTS

RESULT 1

AW669676

LOCUS: 113173 MARC lBOV Bos taurus cDNA 5', mRNA

EST 25-APR-2001

DEFINITION: AW669676

ACCESSION: AW669676.1

VERSION: GI:7526190

KEYWORDS: EST.

SOURCE: COW.

ORGANISM: Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

REFERENCE: 1 (bases 1 to 149)

AUTHORS: Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W., Rohrer,G.A., Chitko-McKown,C.G., Perera,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

JOURNAL

MEDLINE

COMMENT

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smitht@mail.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -mnscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTTCCCACTCAGCAGC

Plate: 107 row: D column: 22



Seq primer: ATTAGGTGACACTATAG.  
Location/Qualifiers  
1..149

## FEATURES

source

/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC IBOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;  
Library made from pooled tissue from lymph node, ovary,  
fat, hypothalamus, and pituitary."  
24 a 48 c 38 g 39 t

BASE COUNT  
ORIGIN

Query Match 3.4%; Score 87; DB 10; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2459 ctgtacctcttagtgcagccatctgtttgtccctcccccgtgccttcttgacc 2518

Db 36 CTGTGCTTCTAGTTCAGCCACCATCTGTGTTCCTCCCTCCCGTGCCTTCTTGACCC 95

QY 2519 tggaaagtgcacccactgccttt 2545

Db 96 TGGAGGTGCCACTCCCACTGCTCTT 122

## RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 2459

Db 36

QY 2519

Db 96

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps



Db	109	TGGAAGTGGCCACTCCCACTGTCCTTT	135
RESULT	4		
LOCUS	AW785901	162 bp	mRNA
DEFINITION	117399 MARC iPig Sus scrofa cDNA 5', mRNA sequence.	EST	09-JUL-2000
ACCESSION	AW785901		
VERSION	AW785901.1	GI:7842677	
KEYWORDS	EST.		
SOURCE	pig.		
ORGANISM	Sus scrofa		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
AUTHORS	1 (bases 1 to 162) Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, J.W., Laegreid, W.W., and Keele, J.W.		
TITLE	Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smithemail.marc.usda.gov Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options. PCR Primers FORWARD: AGGAACAGCTATGACCAT BACKWARD: GTTTTCCAGTCACGACG Plate: 37 row: K column: 23 Seq primer: ATTTAGGTGACACTATAG. Location/Qualifiers 1..162 /organism="Sus scrofa" /db_xref="taxon:9823" /clone_lib="MARC LPIC" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."		
FEATURES	Source		
BASE COUNT	28 a	51 c	42 g
ORIGIN	41 t		
Query Match	3.4%;	Score 87;	DB 10; Length 162;
Best Local Similarity	100.0%;	Pred. No. 1.3e-07;	
Matches	87;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	2459	ctgtgctctctagtgcagacatctgtgttggccctcccccgccttccttgacc	2518
Db	49	CTGTGCTCTCTAGTGGCCAGCATCTGTTGGCCCTCCCGTGCTTCCTTGACCC	108
QY	2519	tggaaggtgccactccactgtccttt	2545
Db	109	TGGAAGTGGCCACTCCCACTGTCCTTT	135
RESULT	5		
LOCUS	AW669518	243 bp	mRNA
DEFINITION	112921 MARC IBOV Bos taurus cDNA 5', mRNA sequence.	EST	25-APR-2001
ACCESSION	AW669518		
VERSION	AW669518.1	GI:7526032	
KEYWORDS	EST.		
SOURCE	cow.		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.		
AUTHORS	1 (bases 1 to 503) Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G., Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.		
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle		



JOURNAL  
MEDLINE  
COMMENT

Genome Res. 11 (4), 626-630 (2001)  
21180013  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.

PCR Primers  
FORWARD: AGGAACACGATGACCAT  
BACKWARD: GTTTCCAGTCACGAGC  
Plate: 82 row: E column: 4  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1..503

FEATURES  
source

/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 2BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from testis, thymus,  
semitendinosus muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."  
BASE COUNT 122 a 146 c 130 g 105 t  
ORIGIN

Query Match 3.4%; Score 87; DB 11; Length 503;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2459 ctgtgccttctagtgcagccatctgtttgccccctccccctgcttctcttgacc 2518  
|||||  
Db 367 CTGTGCTTCTAGTTGCCAGCCATCTGTTGTGCCCTCCCGCTCCCTTCCCTTGACCC 426  
|||||  
QY 2519 tggaggtgcccactcccactgtccttt 2545  
|||||  
Db 427 TGGAGGTGCCACTCCCACGTGCTTT 453  
|||||

RESULT 7  
AW669634/c

LOCUS  
DEFINITION 113105 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.  
ACCESSION AW669634  
VERSION AW669634.1 GI:7526148  
KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus

REFERENCE  
AUTHORS  
1 (bases 1 to 328)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett  
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and  
Keele,J.W.

## TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred

v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.

PCR Primers  
FORWARD: AGGAACACGATGACCAT  
BACKWARD: GTTTCCAGTCACGAGC  
Plate: 107 row: I column: 14  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1..328

FEATURES  
source

/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from lymph node, ovary,  
fat, hypothalamus, and pituitary."  
BASE COUNT 71 a 101 g 68 t  
ORIGIN

Query Match 3.4%; Score 85.4; DB 10; Length 328;  
Best Local Similarity 98.9%; Pred. No. 2.4e-07;  
Matches 86; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2459 ctgtgccttctagtgcagccatctgtttgccccctccccctgcttctcttgacc 2518  
|||||  
Db 110 CTGTGCTTCTAGTTGCCAGCCATCTGTTGTGCCCTCCCGCTCCCTTCCCTTGACCC 51  
|||||  
QY 2519 tggaggtgcccactcccactgtccttt 2545  
|||||  
Db 50 TGGAGGTGCCACTCCCACGTGCTTT 24  
|||||

RESULT 8  
AW669633/c  
LOCUS  
DEFINITION 113104 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.  
ACCESSION AW669633  
VERSION AW669633.1 GI:7526147  
KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus

REFERENCE  
AUTHORS  
1 (bases 1 to 374)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett  
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and  
Keele,J.W.

## TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.

## PCR Primers

FORWARD: AGGAACACGATGACCAT  
BACKWARD: GTTTCCAGTCACGAGC  
Plate: 107 row: I column: 13  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1..374

FEATURES  
source

/organism="Bos taurus"



```

/db_xref="taxon:9913"
/clone_lib="MARC lBOV"
/tissue_type="pooled"
/lab_host="DH108"
/Note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
79 t
BASE COUNT      84 a 102 c 109 g
ORIGIN

Query Match      3.3%; Score 84.4; DB 10; Length 374;
Best Local Similarity 98.8%; Pred. No. 3.8e-07;
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2459 cttgctcttagtcagccatctgtttgcttgcctcccccgtgcttcttgcaccc 2518
Db 109 CTGTGCTCTTAGTTCAGCCAGCATCTGTGTTGGCCCTCCCGGTCCTTCTTGACCC 50

QY 2519 tggaagtgccactccactgctctt 2544
Db 49 TGAAGGTGCCACTCCCACTGTCTT 24

RESULT 9
BG365021      98 bp mRNA EST 25-APR-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 98)
Smith, P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrrenkrug, S.C., Bennett
, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perteau, G., Holt, I., Karameycheva, S., Liang, F., Quackenbush, J. and
Keefe, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCACGAGC
Plate: 72 row: L column: 16
Seq primer: ATTTAGTGACACTATAG.
Location/Qualifiers
1..98
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC lBOV"
/tissue_type="pooled"
/lab_host="DH108"
/Note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
27 t
BASE COUNT      18 a 35 c 18 g
ORIGIN

Query Match      2.4%; Score 61.2; DB 13; Length 1001;
Best Local Similarity 25.4%; Pred. No. 0.017;
Matches 87; Conservative 106; Mismatches 149; Indels 0; Gaps 0;

QY 599 gggtaatttataaagtttgggtgcgcgaagcaattgctgctaactatgacctaaag 658
Db 391 GAAGRTTAAARKKTKTKTKTSATKKCAARSYTTYGRAAWKSTYTWMTWTWMMGG 332

QY 659 aggcataagtttacaagattttccgctctatatgtatggtgcacacag 718
Db 331 GGSWTKGGGTGGRWKGAACCCGRARGAWTTTCCCVCCWWRCCMWKGCCTKGGRRCS 272

QY 719 attctatcattcagagcttatggggttactccaaaccagcaaacgataaactcgga 778

```

```

Query Match      2.8%; Score 71; DB 11; Length 98;
Best Local Similarity 100.0%; Pred. No. 0.00025;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2475 ccagccatctgtttgcccctccccctgcttcttgcacccgagaggtgccaactcc 2534
Db 1 CCAGCCATCTGTGTTGCCCTCCCGCTTCTTGACCTGGAAGGTGCCACTCC 60

QY 2535 cactgtccttt 2545
Db 61 CACTGTCTCTT 71

RESULT 10
CNS007BE/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1001)
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
RP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..1001
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR15H24"
/Note="end : TET3"
BASE COUNT      159 a 148 c 151 g 212 t 331 others
ORIGIN

Query Match      2.4%; Score 61.2; DB 13; Length 1001;
Best Local Similarity 25.4%; Pred. No. 0.017;
Matches 87; Conservative 106; Mismatches 149; Indels 0; Gaps 0;

QY 599 gggtaatttataaagtttgggtgcgcgaagcaattgctgctaactatgacctaaag 658
Db 391 GAAGRTTAAARKKTKTKTKTSATKKCAARSYTTYGRAAWKSTYTWMTWTWMMGG 332

QY 659 aggcataagtttacaagattttccgctctatatgtatggtgcacacag 718
Db 331 GGSWTKGGGTGGRWKGAACCCGRARGAWTTTCCCVCCWWRCCMWKGCCTKGGRRCS 272

QY 719 attctatcattcagagcttatggggttactccaaaccagcaaacgataaactcgga 778

```



```

Db 271 GRWTTTWTATTTTWWTTTWWTTTACCTTAANKGSGGKTWRWKGKKAATTTWRCCOC 212
Qy 779 gtgatttacttcgaagttgtaggtatataatttcacgtttaaagtcacatttt 838
Db 211 MAARKTWASWTTTGGTGTWAWTATKGTWAWKGTGKGYCYWAAWYKTTTK 152
Qy 839 aataaactcttaaacacagcgtcgacattattatttagtagagcttttttttttt 898
Db 151 RAKKSATWTTWWAARRRRGGGRRGAAATTTTATWWSGTWTTTWTTRAAWW 92
Qy 899 tataataaactaaagattttattatttttttttttttttttttttttttttt 940
Db 91 WTTAMKRAAARSMTTKRAAKRAWTAWTTTTRAWTTWTGA 50

RESULT 11
CNS0039G/c 1101 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE Direct Submission
AUTHORS Aaron Mamoser in Pieter de Jong's laboratory in the Department of
TITLE Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
JOURNAL NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and Est libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
    Location/Qualifiers
        source
            1..1101
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone_lib="RPCI-98"
                /clone="BACR08K10"
                /note="end : TET3"
BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match 2.4%; Score 60; DB 13; Length 1101;
Best Local Similarity 17.3%; Pred. No. 0.03;
Matches 119; Conservative 269; Mismatches 298; Indels 0; Gaps 0;

Qy 241 gtgaattgtagttgtagtggaacacattgcttcttcagaacacactccactacagaa 300
Db 1101 GKARRGGDDTWRDTRKDDDDWTKWTTWTKDRADRRWAGDADRWWDGAGTWWTATW 1042
Qy 301 aaagctgctacaaacgctatgaatacaataactgtgtttggcagtgctgctcggaag 360
Db 1041 WWWATWTDWTKWKKWATAAKTDTATWTTWRTAWRADWAGDRGAKRDRDAATDAG 982

```

```

Qy 361 catagtcagggttccttgatcaatggacagaaaaaacctctctatatcttattgagcttcc 420
Db 981 AGRRDGGRRKDKDKDKGDDDDKGGKKKKAACAATWATKWDDWDKWDKWDGAKDRK 922
Qy 421 ttaatgaaccccttttagcaaggtacgaagactagcttaaatggaaaaaaccttagct 480
Db 921 ADDDGGAGDKDDGKGKADDDDTGCTKDDDDKDKWDDKAGTWDGATWAAATDWWWW 862
Qy 481 tgggtattgagggaacttttaggggactcagaaaaagctggagactgtgtgccacgta 540
Db 861 GWADADWTTWDAADDDWADDDWADWAWKWDADAWAGARTADRRDWDGDRACKRGARKRR 802
Qy 541 cgttatgagtatgctgaagccttgctcagttccagaaaatagatgtttcaggagattggccgt 600
Db 801 DKRADDKRDADDDDDAATTTTWTTRDTRDDDKWKTDTWTRWAADRTWDRDDDDDRD 742
Qy 601 ggtaattattaaagtttgggtgcgccaagaactgtctcaactatgatcctcaagaag 660
Db 741 RAGTAGRWRRRTWKRKRRRTRWDDADADTARDRRRRGGDDGADAGKGTGKRKRRR 682
Qy 661 gctaagttttacaaattataaaggattttccgcgtctctatatgtatgttggtcacagat 720
Db 681 DRATWDRTDANWADAAMWTTTDTDDTDWDRRRKRGARRRRRTTARAADWMTWKAWDW 622
Qy 721 tctctatcattcagagcttatgggcttactccaaacagcaaacgataaactcggcagt 780
Db 621 AKDWKTRADRWDAADTWDARADRWDAKARAWRARRDRARARADRRTWTGKTTT 562
Qy 781 gatttactctccgaagtttgatcaggtatataatttcagcgcttttaagtcacaaatttaa 840
Db 561 ATWTTWAARAAMWAWATTTATWTTTWTWTTTWTWTTTWTWTTTAAWAAWATAT 502
Qy 841 taaaacttttaaacagggtgcgcattatatttagtgagagcttttttttttttttttt 900
Db 501 WAAWTAATAAATAAATAAATTTTWTWTAATAAATAAATAAATAAATAAATAAATAA 442
Qy 901 taataaaactaaagatttttttttttttttttttttttttttttttttttttttttt 926
Db 441 TTTTWTWTTWTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 416

RESULT 12
CNS00F0UH/c 996 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR31021 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL071063
VERSION AL071063.1 GI:4951105
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 996)
REFERENCE Direct Submission
AUTHORS Aaron Mamoser in Pieter de Jong's laboratory in the Department of
TITLE Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
JOURNAL NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's

```



P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

1..996

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="RPCI-98"

/clone="BACR31021"

/note="end : TET3"

BASE COUNT 383 a 164 c 81 g 171 t 197 others

## ORIGIN

## Query Match

Best Local Similarity 38.1%; Score 59.4; DB 13; Length 996;

Matches 111; Conservative 48; Mismatches 132; Indels 0; Gaps 0;

QY 664 aatagtttacaattataaaggatttccgctctatatatgtatggcatcacagattct 723

Db 995 AAAATTTTATATAAAAAAATTTTWTWATTWMAAAATWAWAWATWMAATTTWTA 936

QY 724 ctatcattcagagcttatgggcttactccaacagcaaacgataaaactcggcagat 783

Db 935 ATTBTFTBGGSSGSSGSSGRRRAAAAAAAMWMAATTTTWTWAAW 876

QY 784 ttacttccgaagttgatctaggtataattcagcgttttaagtcacaaatttaataa 843

Db 875 TTTAARTTWRATTWATTAATTAATTTTATTTTATTTTATTAATTAATTTTATTTT 816

QY 844 aatctttaaacaagcgcgcgaattatttagtgagagcgtttttttttttttataa 903

Db 815 ATTATATTTTAAWAAWTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 756

QY 904 taaactaaagattttatt 954

Db 755 TWAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 705

## RESULT 13

CNS014PQ/c

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence SP6 end of BAC BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

## FEATURES

source

1..987

/organism="Drosophila melanogaster"

/plasmid="pBeloBAC11"

/db\_xref="taxon:7227"

/clone\_lib="DrosBAC"

/clone="BACN12P22"

/note="end : SP6"

BASE COUNT 257 a 122 c 122 g 241 t 245 others

## ORIGIN

## Query Match

Best Local Similarity 32.4%; Score 56; DB 13; Length 987;

Matches 91; Conservative 65; Mismatches 125; Indels 0; Gaps 0;

QY 663 taatagtttacaattataaaggatttccgctctatatatgtatggcatcacagattc 722

Db 911 WWHHTTTHHWWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTT 852

QY 723 tctatcattcagagcttatggggttactccaacagcaaacgataaaactcggcagtg 782

Db 851 TATAMWAAAAAAMWMAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTA 792

QY 783 ttttacttccgaagttgatctaggtataattcagcgtttttaaagtcacaaatttaata 842

Db 791 ATTAATWAAWMAAAATTTWAATTTTAATTAATTTTCTTAAAWWATWTTAAWATTTT 732

QY 843 aatctttaaacaagcgcgcgaattatttagtgagagcgtttttttttttttata 902

Db 731 WWTTTTWTAAAAAAMWMAATTTTAAWMAATTTTAAWMAATTTTAAWMAATTTTAAW 672

QY 903 ataaactaaagatt 943

Db 671 TTTTATTCTTAATTTTGTWTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 631

## RESULT 14

CNS04E5M

LOCUS

DEFINITION

Tetraodon nigroviridis genome survey sequence T7 end of clone 103P02 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:22:17 ; Search time 12230.8 Seconds  
(without alignments)  
878.082 Million cell updates/sec

Title: US-09-391-606-2  
Perfect score: 651  
Sequence: 1 atgacaaaaaacattatgc.....taggtataatttcagcgcttt 651

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_om.\*  
20: em\_or.\*  
21: em\_ov.\*  
22: em\_pat.\*  
23: em\_ph.\*  
24: em\_pl.\*  
25: em\_ro.\*  
26: em\_sts.\*  
27: em\_sy.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htgo\_hum.\*  
31: em\_htgo\_inv.\*  
32: em\_htgo\_rod.\*  
33: em\_htg\_hum.\*  
34: em\_htg\_inv.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	647.8	99.5	2238	6	AX045137	Sequence
2	647.8	99.5	10421	1	AE001654	AE001654 Chlamydia
3	647.8	99.5	12548	1	AE002165	AE002165 Chlamydia
4	647.8	99.5	300550	1	AP002547	AP002547 Chlamydia
5	643.8	98.9	2148	1	CHT768DA	L23921 Chlamydia p
6	240.8	37.0	10954	1	AE001333	AE001333 Chlamydia
7	237.	36.4	14698	1	AE002357	AE002357 Chlamydia
8	201.4	30.9	3455	1	CTU50732	U50732 Chlamydia t
9	46.2	7.1	143364	2	AC015625	AC015625 Homo sapi
10	44.8	6.9	1141	6	AX083744	AX083744 Sequence
11	40.4	6.2	347550	1	AP001118	AP001118 Buchnera
12	40.4	6.2	174327	9	AC007091	AC007091 Homo sapi
13	39.8	6.1	175384	9	AL451103	AL451103 Human DNA
14	38.8	6.0	179033	2	AC023295	AC023295 Homo sapi
15	38.4	5.9	161888	9	HS70501	AL035251 Human DNA
16	37.6	5.8	135978	9	CNS01DV8	AL133485 Human chr
17	37.6	5.8	151017	9	AL591431	AL591431 Human DNA
18	37.6	5.8	191736	33	AC025411	AC025411 Homo sapi
19	37.6	5.8	193594	2	AL592563	AL592563 Homo sapi
20	37.4	5.7	41718	3	CELFI6G10	AF077537 Caenorhab
21	37.4	5.7	62622	2	AC068814	AC068814 Homo sapi
22	37.4	5.7	144165	9	AC002127	AC002127 Human BAC
23	37.4	5.7	228999	2	AC083885	AC083885 Homo sapi
24	37	5.7	159027	2	AC016647	AC016647 Homo sapi
25	37	5.7	159611	9	AC026475	AC026475 Homo sapi
26	37	5.7	169195	2	AC090213	AP090213 Homo sapi
27	37	5.7	185754	2	AP002406	AP002406 Homo sapi
28	37	5.7	185754	2	AP002406	AP002406 Homo sapi
29	37	5.7	188072	2	AC009040	AC009040 Homo sapi
30	36.8	5.7	157000	8	AC007627	AC007627 Genomic S
31	36.6	5.6	13933	14	AF013255	AF013255 Human res
32	36.6	5.6	15225	6	AR089138	AR089138 Sequence
33	36.6	5.6	15225	14	AF013254	AF013254 Human res
34	36.6	5.6	124321	9	HS417615	AL009174 Human DNA
35	36.6	5.6	162084	2	AP001093	AP001093 Homo sapi
36	36.4	5.6	63747	3	AC025727	AC025727 Caenorhab
37	36.4	5.6	161296	2	AL138963	AL138963 Homo sapi
38	36.4	5.6	168050	9	AC090526	AC090526 Homo sapi
39	36.4	5.6	173691	33	AC009823	AC009823 Homo sapi
40	36.4	5.6	209999	2	AC019128	AC019128 Homo sapi
41	36.2	5.6	3519	8	ATH297264	AJ297264 Arabidops
42	36.2	5.6	91566	8	F25C20	AC007296 Arabidops
43	36.2	5.6	110000	2	AC073702_1	Continuation (2 of
44	36.2	5.6	154120	2	AC055114	AC055114 Homo sapi
45	36.2	5.6	157016	2	AC021866	AC021866 Homo sapi

## ALIGNMENTS

RESULT	1	AX045137	2238 bp	DNA	PAT	24-NOV-2000
LOCUS		Sequence	7 from Patent	WO0066739.		
DEFINITION		AX045137				
ACCESSION		AX045137.1	GI:11343736			
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
Chlamydia pneumoniae.						
Chlamydia pneumoniae						
Bacteria; Chlamydiales;						
Chlamydiae; Chlamydia						
1 (bases 1 to 2238)						
REFERENCE						
Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.						
1 (Chlamydia) antigens and corresponding dna fragments and uses thereof						
PATENT: WO 0066739-A 7 09-NOV-2000;						
Aventis Pasteur Limited (CA)						
Location/Qualifiers						
1..2238						
/organism="Chlamydia pneumoniae"						
/db_xref="taxon:83558"						

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES



CDS 766..2238  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC17330.1"  
 /db\_xref="GI:11343737"  
 /translation="MVNPIGPGPIDETERTPPADLSAGGLEASAAKSAEQAQRTAGAE  
 AKPKESKTSVERWSILRSVAVNALMSLADKLGIASSSSSTSRSDVDTATATPT  
 PPPTFDYKTAQATYDTIFTSTLADIAQALVSLQDAVNIKDTAATDEETATAAEW  
 ETKNADAVKGAQITELAKYASDNOAILDSGLKTSFLLQALLOSVAANNKAAELL  
 KEMQDNPVPGKTPAIQAQSLVDQDATQIEKDGNAIRDAYFAGQNASGAVENAKSN  
 NTSNIDSAKAATIAQTAQIAEQKFPDPSILQEAQOMVIAQEKDKLNKIPAGSDV  
 PNPTTGVGSKQSGSSIGSRVSMLLDDAENETASILMSGFROMIHMFENTENPDQAA  
 QOELAAQARAAGDSDSAAALADAKALEALGKAGQOQGLNALGQIASAAVWSA  
 GVLPQQVLTIRARYQAYVEOKLISEDLNSAVDHHHHH"  
 BASE COUNT 689 a 459 c 497 g 593 t  
 ORIGIN  
 Query Match 99.5%; Score 647.8; DB 6; Length 2238;  
 Best Local Similarity 99.7%; Pred. No. 5.6e-155;  
 Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 atgacaaaaaacattatgcttggttagagaggtctcaatcgcttgcctaaacag 60  
 Db 1 ATGACAAAAAACATTATGCTTGGGTGTGAGAGGGATTCTCAATCGTTTGCCTAAACAG 60  
 QY 61 tttttgtgaatgtagttgtcgtggaacacattcgttcttcagaaacctccact 120  
 Db 61 TTTTGTGGAATGTAGTGTGTCGACGTGGACACATTCGTTCCTTCAGAACCTCCACT 120  
 QY 121 acagaaaaagctgctacaaacgctatgaaatacaataactgtgttggcagtggtcgtc 180  
 Db 121 ACAGAAAAAGCTGCTACAAACGCTATGAAATACAAATAGTGTGTTGGCAGTGGCTCCTC 180  
 QY 181 gaaagcatagtcagttccttggtatcaatggacagaaagcctctatatctttatgga 240  
 Db 181 GAAAGCATAGTCAGTTCCTTGGATCAATGGACAGAAAAGCCCTCTATATCTTTATGGA 240  
 QY 241 gctttcttaataagacccctttagcaaggtcagaaactacgttaaatggaagaaac 300  
 Db 241 GCTTTCTTAATGAACCCCTTAGCAAGGCTACGAAAGACTACGTTAAATGGAAGAAAC 300  
 QY 301 ctagcttggtttattgaggaacttttagggagctcagaaagctggagactggtcgtcc 360  
 Db 301 CTAGCTTGTTTATTGGAGAACTTTAGGGGACTCAGAAAAGCTGGAGACTGCTGCC 360  
 QY 361 acagtacgttatgattgtcgaagcctgtcagttccagatccagaaatagattttcaggatt 420  
 Db 361 ACAGTACGTTATGATTGTGGAAGCCTGTCGGTTCAGAAATAGATTGTTTCAGGGATT 420  
 QY 421 ggcctggttaatttataaagtttgggttcgcccagcaattgctgctaaactatgatcct 480  
 Db 421 GSCCTGTTAATTATTAAGTTTGTGTCGCCAAGCAATGCTGCTAACTATGATCCT 480  
 QY 481 aaagaggctaattttacaaataataaaggattttcgcgtctatatatgtatggcattc 540  
 Db 481 AAAGAGGCTAATGTTTACAAATATATAAGGATTTTCGGCTCTATATATGTATGGCATC 540  
 QY 541 acagattctctatcattcagacttatgggttactctcaaacagcaaaacgataaacctc 600  
 Db 541 ACAGATTCTCTATCATTAGAGCTTATGGGCTTACTCTCAACACGCAAAACGATAAATC 600  
 QY 601 ggcagtgatttacttctcgaagtttgcattctaggtatataatttcagcgttt 651  
 Db 601 GGCAGTGATTTACTTCCGAAAGTTTGCATCTAGGTATTAATTTACAGCGTTT 651  
 RESULT 2  
 AE001654/c  
 LOCUS  
 DEFINITION Chlamydia pneumoniae section 70 of 103 of the complete genome.  
 ACCESSION AE001654 AE001363

VERSION AE001654.1 GI:4377031  
 KEYWORDS Chlamydia pneumoniae CWL029.  
 SOURCE Chlamydia pneumoniae CWL029.  
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. 1 (bases 1 to 10421)  
 REFERENCE 1 (bases 1 to 10421)  
 AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.  
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
 MEDLINE 99206606  
 PUBMED 10192388  
 REFERENCE 2 (bases 1 to 10421)  
 AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
 FEATURES  
 source  
 1..10421  
 /organism="Chlamydia pneumoniae CWL029"  
 /strain="CWL029"  
 /db\_xref="taxon:115713"  
 133..2751  
 /gene="CpN0727"  
 133..2751  
 /gene="CpN0727"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CT619 hypothetical protein"  
 /protein\_id="AADI8866.1"  
 /db\_xref="GI:4377032"  
 /translation="MSTFSTQNRLRTISGSTRITKLDHKYSGFDPSPVAINLEELN  
 SGYALRHLNALQSENTVAALLNPNTIFPTTSWTDYKHSRQASPRAPSSQTP  
 DIVSAAALALVLDGGLVDELVAIDTIDGALSTISVROLMASLGLTTLTAQEK  
 VFFSSVYPSEKLEHVKQEAIEIAQKEEIKAVLEAKGVSTEEIEAILKEYPDII  
 AADFEKFEIEPLHYRAKVGAPTEQENNAIQLPTPPAITPDNVNENGMTLSI  
 LQATDDAIKAQALGQDEIITLQTLVPLVDKTTKAEFDLYTATQLPNTASLKL  
 YLTDQIAEYRGKITVYQNSIQNSLSETRKVVNRRSMLETQSMFOQACNCFVTWLS  
 QANALNIAITNKYISAVLTITSMEMVGGILCLISYMYERLADDEKALFDSVNEYLIPIHI  
 VVGSWVNGHIAKMAAQOELAEVSLGTAVTSQDQIKAYLQTRGNEPKATRHFFINIG  
 QMVFANETVGNCLTTANGAIPDGLGFIEMANTVGTVEADIVSNRQALNENFNTA  
 ATAHVQLQIAELQKADDDLPKASFTENRKFVAVAMITSESGLDLSMLNSQ  
 LPQEAFLKELIEINFLNLAANALNSLQITNEFTSTVSYSLSYVOSKQNL  
 AGDYETLLAAAREYIYRDARCKQALINLVNGLLQKINSLPGATSAQKQMLNAT  
 YQYSLVTLNQLTVLESLLAGLKWLTQTTNNKYDKSVFKIESFDWIPTLAALESF  
 LTSGFNLSATGGIGLPFTQVQSDQTYTSQGTQQLNLQNMTHIQEWTLVSTSMQ  
 VLNGILSQLAGAIYSN"  
 complement(3078..5033)  
 /gene="CpN0728"  
 complement(3078..5033)  
 /gene="CpN0728"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CHUPN 76 kDa Homolog\_1 (CT622)"  
 /protein\_id="AADI8867.1"  
 /db\_xref="GI:4377033"  
 /translation="MVNPIGPGPIDETERTPPADLSAGGLEASAAKSAEQAQRTAGAE  
 AKPKESKTSVERWSILRSVAVNALMSLADKLGIASSSSSTSRSDVDTATATPT  
 PPPTFDYKTAQATYDTIFTSTLADIAQALVSLQDAVNIKDTAATDEETATAAEW  
 ETKNADAVKGAQITELAKYASDNOAILDSGLKTSFLLQALLOSVAANNKAAELL  
 KEMQDNPVPGKTPAIQAQSLVDQDATQIEKDGNAIRDAYFAGQNASGAVENAKSN  
 NTSNIDSAKAATIAQTAQIAEQKFPDPSILQEAQOMVIAQEKDKLNKIPAGSDV  
 PNPTTGVGSKQSGSSIGSRVSMLLDDAENETASILMSGFROMIHMFENTENPDQAA  
 QOELAAQARAAGDSDSAAALADAKALEALGKAGQOQGLNALGQIASAAVWSA  
 GVLPAAASIGSSVQKLYKTSKSGDKTQISAGYDAYKINSDAYGRANDTRDVI  
 NVSTPALTRISVPRATEARPEKTDQALARLTSNRSRTGLDVTYSQVSAQVNMIIQ  
 SNFOANNEIRKLTSAVTKPPQFGYVQLNSDSTQKFTAKLESLEFAEGSRATAEIK  
 ALSFETNSLFTQVVLNIGLSYGLQ"  
 complement(5145..6413)  
 /gene="CpN0729"  
 complement(5145..6413)  
 /gene="CpN0729"



```
/codon_start=1
/transl_table=11
/product="CHLPN 76 kDa Homolog_2 (CT623)"
/protein_id="AADI8868.1"
/db_xref="GI:4377034"
/translation="MKQVYQWLASVLLALTISGYAELPSEQKVKSHYTTTLDVEVK
DYLKRGFVETRKQDGLRIAGDVRLWLYPREDIKNPSDKDKYNPLPVRNRYSEFYI
YIDYAEENWLSKNNMTAAGENTAAGVDINRAFLCYREYKNPEYRTDFPFMVEIGRS
GLDGLFEQVQSNFGLHLYWTRELSKDYQYOVIVHGGPFVYVNMTKKHVAWVEGI
LNRLPKQFVKYCSVDWNTFVPSSTTEKAATNAMKYKVCVWQWMLKKHQSQVPIWG
QKPLYLYGAFMLPLAKATKTLNGKENLAWFVGGTGLGLRKAAGDSATVRYEYVEA
LSVPEIDVSGLRGNLKLWFPAQAIANIDPEANGFTNYKGFSAIYMGITDLSLSPR
AYSGPANDKLGSDFTFRKDLGIISAF"
gene
complement(6516..8159)
/ gene="mvln"
CDS
complement(6516..8159)
/ note="CpN0730"
/ gene="mvln"
complement(6516..8159)
/ gene="mvln"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AADI8869.1"
/db_xref="GI:4377035"
/translation="MSRKDNVSLARSIFNILLSTGFCRSRTIGFRETAMATYFGADPI
VAEWLGRVFFLRKLLGLLEQAFIPEEFRAQSLDRAAEFFRRFRSLIKGSTI
IFTLLIEAVLMVLOYVEEGYDMLITMLLPCGIFIMMYNVNGLHCHENKFFGVG
LAPVYNIITFFIARHSDPRERIIGLSVALVIGFFEWLITVPGVWKPLEAKSP
POEHSVRALLASLGILTSIFQLNLLSDICLARKVHEIGPLYLMYSLKIYOLPHT
LFGFVFTVLLPATSRVOREDERHGLKMLFVLTLMSVMIINTAGLLIALLPVGRV
LYEHGLPQSAVAIVRVLYSGYASIIIPMALAPLVLYVLYAQRQYVPLFGIGITALA
NIVGLTLGRWLKDVSGISATSIITAWQLPLFYLYSSKRLPMYSLVLLMSIRRSIK
NVTMTLACMLTGLNLTQTYIPLNPLTFLAWPLSSITPAQIAIFLSESCIFLAFI
FGFAKLRLVEDLINLASFEYWRGQRLQRLQHVMDQTON"
8143..8319
/ gene="CpN0731"
8143..8319
/ gene="CpN0731"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8870.1"
/db_xref="GI:4377036"
/translation="MSFLINKFPFPPPLKQDYLFDISPTSSSEITIGGSYFKLNKASL
QSSILRLRSIIIS"
8654..9535
/ gene="nfo"
8654..9535
/ note="CpN0732"
8654..9535
/ gene="nfo"
/codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AADI8871.1"
/db_xref="GI:4377037"
/translation="MKVLPPSIPLLGAHTSTAGLKNALIVEGRDICASTVQIETANO
ROMORRALKEVIEDFKAALKETDLSYIMSHAGYLINPGADPDVILKSRIGIYOEIL
DCITIGSFVNFHGAALKSKEDCMNKSIFSQSAPLFSDSPPLVLLVETTAGQGT
LIGNESELGYLVONLKNQIPIGVCDTCHIFAAGYDITSPQWEDVINEFVEYGLS
YLRFLHNSMFPILGANKDRHAPLGEYIGKESFKFLMTDERTKIPKYLETPGPGEN
WQKEIGELKLFKNRDS"
complement(9654..10283)
/ gene="rs4"
/ note="CpN0733"
complement(9654..10283)
/ gene="rs4"
/codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AADI8872.1"
/db_xref="GI:4377038"
/translation="MARYCGPKNRVARRFGANIFGRSNRPLKPPHPPGQHMQRKK
SDYGLQLEEKQKLKACYCMIMEKOLVRAKVEYIHKQGNVAOMFLPERCECRLDNMVYRM
GFAKTIFAAQQLVAHGHLVNGRRVDRRSFFLRPGMQISLKEKSKRQSVKDALESKD
```

```
BASE COUNT 3278 a 2158 c 1983 g 3002 t
ORIGIN
ESSLPYSISLDKTKGKGLLVSPQDQIEAQLPLPINISVVCEFLSHRT"
Query Match 99.5%; Score 647.8; DB 1; Length 10421;
Best Local Similarity 99.7%; Pred. No. 5.8e-155;
Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 atgacacacacacacattatgcttgggttttagaagggaattctcaatcgcttgcctaaacag 60
5798 atgacacacacacacattatgcttgggttttagaagggaattctcaatcgcttgcctaaacag 5739
61 tttttgtgaaatgtagtgggttgcacacacattcgcttccctcagaacacccctc 120
5738 tttttgtgaaatgtagtgggttgcacacacattcgcttccctcagaacacccctc 5679
121 acagaaaagctgctacaaacgctatgaaatatacaataactgcttggcagtggtcctc 180
5678 acagaaaagctgctacaaacgctatgaaatatacaataactgcttggcagtggtcctc 5619
181 ggaagacatagtcaggttcccttgatcaatgacagaaaagcctctatatctttatga 240
5618 ggaagacatagtcaggttcccttgatcaatgacagaaaagcctctatatctttatga 5559
241 gcttcttaataagaccccttttagcaaggctacgaagactacgttaaatggaaaagaaac 300
5558 gcttcttaataagaccccttttagcaaggctacgaagactacgttaaatggaaaagaaac 5499
301 ctagcttggtttattggggaacttttaggggaactcagaagaactggagactggtctgcc 360
5498 ctagcttggtttattggggaacttttaggggaactcagaagaactggagactggtctgcc 5439
361 acagtacattatgtagtgcgaaccccttgacgttcagttccagaaatagattgttcaggatt 420
5438 acagtacattatgtagtgcgaaccccttgacgttcagttccagaaatagattgttcaggatt 5379
421 ggccgttgtaattattaaagtttgggttcgcccaagaactgctgctaaactatgacct 480
5378 ggccgttgtaattattaaagtttgggttcgcccaagaactgctgctaaactatgacct 5319
481 aaagaggctaataagtttcaaaaattataaaggatttccgcctctctatatgtatggcctc 540
5318 aaagaggctaataagtttcaaaaattataaaggatttccgcctctctatatgtatggcctc 5259
541 acagattctctatcattcagagcttattgggttcgcccaagaactgctgctaaactatgacct 600
5258 acagattctctatcattcagagcttattgggttcgcccaagaactgctgctaaactatgacct 5199
601 ggccgttgtaattattaaagtttgggttcgcccaagaactgctgctaaactatgacct 651
5198 ggccgttgtaattattaaagtttgggttcgcccaagaactgctgctaaactatgacct 5148
RESULT 3
AE002165 12548 bp DNA BCT 30-MAY-2000
Chlamydia pneumoniae AR39, section 2 of 94 of the complete
genome.
AE002165 AE002161
AE002165.1 GI:7188948
Chlamydia pneumoniae AR39.
Chlamydia pneumoniae AR39
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
1 (bases 1 to 12548)
Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,
Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J.,
McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
```



JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Nucleic Acids Res. 28 (6), 1397-1406 (2000)

20150255

10684935

2 (bases 1 to 12548)

Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,  
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,  
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J.,  
McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.  
Direct Submission

Submitted (01-MAR-2000) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. 12548

/organism="Chlamydia pneumoniae AR39"

/strain="AR39"

/db.xref="taxon:115711"

/note="synonym: Chlamydia pneumoniae AR39"

297..965

/gene="CP0011"

297..965

/gene="CP0011"

/note="similar to SP:P31218 GB:X71492 PID:296947 GB:U00096

PID:I736770; identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="uridine kinase"

/protein\_id="AAF37907.1"

/db.xref="GI:7188950"

/translation="MLMLMMIIGITGGGAGKTTLTQNIKEIFGEDVSVICQDNYK

DRSHYTPERRANLWDHDFADNDLISDIKRLKNEIVQAPVDFVLGNRSKTEIET

IYPSKVLVEGILVFENQELRLMDIRIFDVTDADEIRLRMVRDVGQEGVDCIMS

RYLSMVKPMHEKTEPTRKYADIIVHGNVRQNVVTNLSQIKHNLENALESDETVM

VNSK"

complement(1047..2021)

/gene="CP0012"

complement(1047..2021)

/gene="CP0012"

/note="conserved hypothetical protein; identified by

Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAF37908.1"

/db.xref="GI:7188951"

/translation="MEKKYALAYVYITRVNDPHEETALHKKFLEDLDVSCRIVISEQ

GINQFSGSYSPHAELYNQWLKERPNFSKIKFKTHIKENIFPRITVYKRELALGCE

VDLSQAKHLSPOEWHKQLKQENKCLILDVRNVEWKIGFDNATLPDIQTFREFPEYA

EKLAECDPTTPMYCTGIGICELYSPVLEKGEVYQLDGGVYAYGOQVGTGKW

LGLFVFDRLAIPIDESDPDAPVIAECCHCQTPSDAYNCANTDCNALFLCDEGCIH

QHGCCGECSSQSPRVKFDSSRGKPFRAHLCEISENSASCCLI"

2238..2867

/gene="CP0013"

2238..2867

/gene="CP0013"

/note="similar to SP:P81288; identified by sequence

similarity; putative"

/codon\_start=1

/transl\_table=11

/product="ribosomal protein S4"

/protein\_id="AAF37909.1"

/db.xref="GI:7188952"

/translation="MARYCGPKNRRVRRFGANIFGRSRNPLKKPPHPPGQHGMRKKK

SDYGLQEEKQVLKACYGMTMEKQVAKFVTHKQNVACMFLEPERCEKLDNMYRM

GFAKTIFAQOLVAHGHLVNGRVDERSFELRPGQISLKEKSKLQSKYDALESKD

ESSLPSVLSIDKGFKGELLYSPFQDQIEAQLPLPINISVVEFLSHRT"

complement(2986..3867)

/gene="CP0014"

complement(2986..3867)

/gene="CP0014"

/note="similar to SP:P12638 GB:M22591 PID:146954

PID:405898 GB:U00096; identified by sequence similarity;

putative"

/codon\_start=1  
/transl\_table=11  
/product="endonuclease IV"  
/protein\_id="AAF37910.1"  
/db.xref="GI:7188953"  
/translation="MKVLPSPPLIGLAHTSTAGLKNAYIEGRDCASTVQIIFRANQ  
RQWRALKEEVEDKAALKETDLSYIMSHAGYLINPAPDPSPIVLEKSGIYQOETL  
DCITLIGTSFVNFHFGAALKSSKDCMNKIVSSQSAPLFDSPPLVVLLETTAGQCT  
LIGNSFEELGYLVONLKNQIPIGVCVDTCHIFAAGYDITSPOGWEDVLNEFDEVGILS  
YLRAFHLNDSMFLGANKDRHAPLGEYIGKESFKFLMDTERTRKIPKYLETTPGGEN  
WQKEIGELLKFSKNRDS"  
complement(4202..4378)  
/gene="CP0015"  
complement(4202..4378)  
/gene="CP0015"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF37911.1"  
/db.xref="GI:7188954"  
/translation="MSFLLIKPFSPPLPKQDYLFDSPTSSEITGGSYFKLNKASL  
QSTLRLRSISIIIS"  
4362..6005  
/gene="CP0016"  
4362..6005  
/gene="CP0016"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF37912.1"  
/db.xref="GI:7188955"  
/translation="MSRKNEVSLARSIFNLSGTCRSRTIGFIREJAMATYFGADPI  
VAAFGLGRTVFFLRKILGLILEQAFIPHFELRAQSLDRAAFFRFRSLKIGSTI  
IFTLLEAVLWVQVVECTYDMILLTLLPGIFLMVNVGALLHCENKFFGVG  
LAPVYVNIWIFVIVAAHSDPRERIIGLSVALVIGVFFELITVPCWKFLEAKSP  
POHDSVRALLAPLSGLITLSSIFQNLNLSLGLARYVHEIGPLLYMLSKLYOLPHI  
LFGVFTVLLPAISRCVREDHERGLKMKFVLTLMVSMIIMTAGLLLLALPGVRV  
LIEHGLFPQSAVYAVIRLVGASIIIPMALAPLVSVLVAQYAVPLPFIGTALA  
NIVTSLVGLRWIKLDSGISYATSIITAWOLYFYSSKRLPMYKLLMESRRSIK  
VMGTLMACMTLGNLITQTTVIFLNLPLTLPWPLSSITAQIAFLSBSCTFLAFL  
FGAKLLRVEDLINLASFVYRGORGLLQHQVMDQTN"  
6108..7376  
/gene="CP0017"  
6108..7376  
/gene="CP0017"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF37913.1"  
/db.xref="GI:7188956"  
/translation="MKQVQWLASVVLLATISGYAELPLSEQVKSHYTTLDDEVK  
DYLKSGFVETRKQDGLRIAGVRAWRWLYFREDINPSDKDYNPLPVNRYSEFYL  
YIDYRAERNLWSKGMVITAGGTAAGVDDINRAFLGIFRYKNPTRTDFFMEIGHS  
GLGDLFESEVQFSQNFEDGLHIYVETRELSKDYPIOVIVHGGFFVYVNMTRKHAYWVEGI  
LNRLPKQFVVKCVDMNTEVPSTTERKATNNAKMYKVCVMQWLVGKHSQYVWVG  
QKPLLYLGAFLMNLAKATKTTLNKGNLAWFTGGTLGLRKAGDWSATVRYVEVA  
LSVPEIDVSGIGRNLKFWFAQAIANYDPKANGFTNYKGSALYMYGITDLSLSPR  
AYCAYSKPANDKLGSDFTRFKDLGIISAF"  
7488..9443  
/gene="CP0018"  
7488..9443  
/gene="CP0018"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"



```

/protein_id="AAF37914.1"
/db_xref="GI:718957"
/translation="MVNPICGPDIETERTPPADLSAQGLEASAAKSAEQAQIAGAE
AKPSEKTDSEVERWSILRSVNALMSLDKLGIASSNSSTSRSDVSDTATATPP
PPPTFDYKQAOAYDTIETSTSLADQALVSLQADAYTNKDTATDDETTAAEW
ETKNADAVKGAQTELAKYASDNQALDSLGKLTSLDQALQVYANNKAEEL
KEMQNDVPVQPTAIAGSLVDQATQIEKDNQAIADYFAGQNASGAVNAKSN
NSISNDASAKAIAKTQIAEAKFPDPSILQEAQMWIQAEDKNTKPADGSDV
PNPTGTTGGKQOQSSIGSRVSMLLDAENETASILMSGFRQIMHFNTEPNDSQA
QOELAAQARAAGDSDSAAALADQALEAALGKAGQOQILNALGQIASAADVSA
GVPPAASSTIGSSVKOLYKTKSGSDYKTKTOISAGYDAYKSIDNAYGRANDATPDVI
NNVSTPALTSPSPARTPEARPEKTDQALARVISGNSRTILGDVYSQVSAQSMQIIO
SNPANNNEERKQKITSVTPPQRFQIVQLSNDSTQRFIAKLESFLAEGSRVTAIEIK
ALSFETNSLFTQQVLVNLIGSLYGLQ"
complement(9770..12388)
/gene="CP0019"
complement(9770..12388)
/gene="CP0019"
/feature="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37915.1"
/db_xref="GI:718958"
/translation="MSTFSIQNRLRTISGSESTRIKIDHKYSGDPRSPVAINLEELN
SGIYALRLHMLALQSENTNVALENPNNTIPPTTSDYDKHSRPOASSRAPSSQTP
DIVSAAALALVLDVGGIAELVASVTEIDLCALSTISTVROLMASYGLTLTLTAQEK
WVSSVVPSEKNNLEHVQKFAEIOAKOEIKAVLEAKGVSPTEETAILKEYPDLY
AADPEFIELEPLHTYRAKYCAPIQEMNENAIQLPTPPALTDPNVNEVGMNLTSTI
LQADDAIKQALPGQOEIITILQTLVPLVDKTFTRAEFDLITATQPLNTASLKL
YLTDQAEYKRGKTKVYQNSIQNLSRNNRSMLETSQSMFOQAQCTWTIS
QANALNTATNKYISAVLTSMENYGLLCLSYMYERLADDEKAIQKSVNEYLPFIH
VVGSGWGTAKMAAYQELAEYSIGTVAQSQDKAYLQTRGNEFKATRHHFNIG
OMYOFANETVGNCLTTPANGAIOPLGDFIREAMTVGTVEADYSVNAORLNFNTA
ATVFIQLOIQIAELOKADDPKASFTENRKEFAVAANTSESIGDALISMLNSQ
LPKQAEFLKPLIEINFNLAANALNSLIQTNEFSTSYVSSLYVQSKTQNLF

```

```

gene
CDS
Query Match 99.5%; Score 647.8; DB 1; Length 12548;
Best Local Similarity 99.7%; Pred. No. 5.9e-155;
Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 atgacacacacacattatgcttggttgaagagattctcaatcggttcacacag 60
Db 6723 ATGACACAAAAACATTATGCTTGGGTGTGAGAGGATCTCAATCGTTGCTTAACAG 6782
QY 61 ttttttggtaagttagttgctgactggacacacattcgttccttcagaaacccctc 120
Db 6783 TTTTGTGTAATGATGTTGCTGACTGGAACACATTCGTTCTTCAGAAACCTCCACT 6842
QY 121 acagacacacacacacacacacacacacacacacacacacacacacacacac 180
Db 6843 ACAGAAAAAGCTGCTACAAACGCTATGAATACAAATACAAATCTGTGTGGCAGTGCCTC 6902
QY 181 gaaagacatagtcaggttccttgatcaatgagacacacacacacacacacacac 240
Db 6903 GGAAGACATAGTCAGTTCCTTGGATCAATGGACAGAAAAACCCCTCTATATCTTATGGA 6962
QY 241 gctttcttaagaccccttagcaagagctacgagacacacacacacacacacac 300
Db 6963 GCTTCTTAAAGACCCCTTAGCAAGGCTACGAAGACTACGTTAAATGGAAGAAAAAC 7022
QY 301 ctgagcttggtttattgaggaacttaggggactcagacacacacacacacacacac 360
Db 7023 CTAGCTTGCTTTATTGGAGAACTTTAGGGGACTTCAGAAAAAGCTGGAGACTGGTCTGCC 7082
QY 361 acagtcagcttatgagtagtcgaaccccttgctcagttccagataatagatgttcaggat 420
Db 7083 ACAGTACGTTATGAGTATGCGAACCCCTTGCTGGTTCAGAAATAGATGTTTCAGGGATT 7142
QY 421 ggccttggttaattattaaagtttgggttcgccacacacacacacacacacacac 480
Db 7143 GCCCGTGGTAATTTAATTAAGTTTGGTTCGCCCAAGCAATTCGCTTAACATATGATCCT 7202

```

```

QY 481 aaagaggctaagtagttttacacaaattataaagatttttcgctctctatatgtatgcatc 540
Db 7203 AAAGAGGCTAAGTCTTTTACAAATATTAAGGATTTCCGCTCTATATATGTATGGCAATC 7262
QY 541 acagattctctattcattcagagcttatgggcttactccaaacacagcaataaactc 600
Db 7263 ACAGATTCTCTATCATTCAGAGCTTATGGGCTTACTCCAAACCAACGATAAATC 7322
QY 601 ggcagtgatttacttccgaagattgctatagtagtataatttcagcgttt 651
Db 7323 GGCAGTGATTCTTCTCCGAAGATTGATCTAGGTATTAATTTACCGCTTT 7373
RESULT 4
AP002547/c DNA BCF 08-DEC-2000
LOCUS Chlamydophila pneumoniae J138 genomic DNA, complete sequence,
DEFINITION section 3/4.
ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008
VERSION AP002547.2 GI:10176693
KEYWORDS
SOURCE Chlamydophila pneumoniae J138 (strain:J138) DNA.
ORGANISM Chlamydophila pneumoniae J138
REFERENCE 1 (sites) Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)
MEDLINE 20330349
REFERENCE 2 (bases 1 to 300550)
AUTHORS Shirai,M.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.
Nutsunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi
755-8505, Japan (E-mail:shirai@po.cc.yamaguchi-u.ac.jp,
Tel:81-836-22-2227, Fax:81-836-22-2415)
On Sep 15, 2000 this sequence version replaced gi:6172298
gi:6172300 gi:6172396 gi:6172398 gi:8978889
AB033786-AB033787, AB033816-AB033817: Submitted (25-Oct-1999).
FEATURES
source
1..300550
/organism="Chlamydophila pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/feature="synonym:Chlamydia pneumoniae (strain J138)"
complement(21..794)
/gene="dapf"
complement(21..794)
/gene="dapf"
/codon_start=1
/transl_table=11
/product="diaminopimelate epimerase"
/protein_id="BAA98725.1"
/db_xref="GI:8978890"
/translation="MAFYSPSTISKYFIYSGAGNRFLIGETLPEVDVRFVLCQETVVD
GFLYLKPSSCADAOLIIIFNSDGRPTWCGNLCRAIAHLASOKKSDISYSTDSGLYS
GYFYSDRVLVDMTADWRASVHLESPDPLKPVVCIHTGVPHAVVILPEISTDL
SILGFLRYHTFSPDGYNNVNFVQILHCOLRVRTYERGVGEETAACTGALASALVV
SNSYWKRESIQIHTWGGELMTVSQNRGRVYLQGSVTRDL"
complement(763..1338)
/gene="clpP_1"
complement(763..1338)
/gene="clpP_1"
/codon_start=1
/transl_table=11
/product="CLP protease"
/protein_id="BAA98726.1"
/db_xref="GI:8978891"
/translation="MADGEVHKLDIIEKLEARRVFFSEPVTEKSASDAIKKLWYL

```



```

ELKDPGPIVIVINSPPGSDVAGFAVWDQIKMLTSPVTVVTVGLAASMGSVLSLCAAP
GRRATPHSRIMIHQPSIGGPIITGQATDLDIHAREILTKARIIDVYVEATNQPRDII
EKAIDRWMTWANEAKDFGLDGLRPSFNDL"
complement(1358. .2917)
/gene="glya"
complement(1358. .2917)
/gene="glya"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/transl_table="MLKVFEKFKFAIVEIFTKVAVVSVLLHFKFLENASGKGQSLAS
TAYLAADHLNAPPSIGERIIDELKDSRHLKMIASENYSLSVOLAMGNLLTDKYC
EGSPFKFYSCENVDALIEWECVETAKELFAADCACQVPHSGADANLLAYMALITHVC
QGPAYKLGKTYTNELTEBEYTLTKAEMSSCVCLGPSLNSGGHLTHGNVPLNWSKLM
RCFPYDNPOTCEFDYAEIRLAKYKPKVLIAGISYSRRLNFAVLKQIAEOCGSVL
WVDMFADVAGCVFDEENPIYADIVTTTHTKLGRPGGLVLAIRLAEYESTLNKA
CPLMMGGLPHVIAAKTVALKEALSDVKRYAHQVNNARRLAERFLSHGLRLTGTG
DNHMMVIDLGLSGISGKIAEDILSSVGIAVNRNSLPDAIGKWDTSIGIRGTALTTL
GMGIDEMEADVADIIVKLRNLRLSCHVEGSSKKNGKELPEAIAQEARDVRNLLLRFP
LYPIDLEALV"
3107. .3778
/gene="CPj0522"
3107. .3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/transl_table="MTLYGLNOKTARKYQAHYLPILTLPYAKSTPONKRALQFLPO
ATHVILSPSTHLFLSRMTSLSKATLTKTYLICIGESTKERLLSLGQVKYVVAQO
BIAGIPIQLAOLPSSARILYDHSLSARPIRELYNRTFFESYPHYTVQVPRKLKNI
LSKTKKIIFTSPTVRAFAPAKIFPRPEPTYWCQGRWTLQEFQFSSQKVSLLLETGK
SRTSP"
3843. .4175
/gene="CPj0523"
3843. .4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/transl_table="MASSATPGDGTAPSLFPPTATPRYNFKLALFVTIALVWIAL
IATIIAGLCHPLCSPFLTAIPLYFISRYICSHYARNVYIALDVVDPHSHKLDQRS
HSPIFSQR"
4223. .5302
/gene="CPj0524"
4223. .5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/transl_table="MSGSPRTSSQSVLSVVPDRKETAPOKTOFTIAKISTLATLASI
ALGALVAGISITVLGNPFLVFLALITLALSVTYFLYHOMTSKVSSNQVKLEQNEK
PLGRAMQKNVDCYSNEMOFYNNHLNPKFVAIQTDASQFPQFTFLGLRVEIKNQST
GIIFNPVTNLNDTATNLSTILYTLKDKSVWDTCKREGGPAKEDYFPSTEVRV
VKLPNEADQTFNLNLSAEKKSILPTFLGHVCQPKSEELPNQOEYRQALLAYENCL
KAATESHAAIIVALPLFTSVYEVPEEPLPKBGTGYWDNQTAQFCKRALLDAIQNTALR
YQSRSLVILODPENTIESQSRSE"
5642. .6406
/gene="CPj0525"
5642. .6406
/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="CT398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"

```

```

/transl_table="MHDALLSILAIQELDIKIMRLMRVKKHOKELAKVQSLKSDIRR
KVOEKLEMLNKTQRDGENRIQEISEQINKLENOAAVKKMEIDFNAITQEMTTANK
EYRSQHLQSLDMKOAGGEDILQESLASTENSSSVSEIKKIEFISIKKINBEGKAL
LEORTLKHATNPPELLSIYERLLNNKKDRVVPIENRVCSGCHIVLTPQHLENLVRKKD
RLIFCHCSRLYQWESQVNAQENSTAKRRRAAV"
complement(6827. .7816)
/gene="kpsF"
complement(6827. .7816)
/gene="kpsF"
/codon_start=1
/transl_table=11
/product="GutQ/KpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/transl_table="MPSPMISTDVCODILGKOKEAVDFEFOAFQPKPEAMQAEKILGH
SGHVFSYGVKCGCVARKVATLOSLSERALFSPVDLLHGLDGLVSPGDIVCLFSKS
GTQOELDDVPHLKSRAILVAITSMPYSNLAALSDLVILVPSVAHLDPNLIPTNST
TCOMIGDFELMILLHSGVSLSTYKGNHPSQGVGMKANGKVKDFNFPKTEVPFCHLG
DKVSFELEFSAYCGCVICVDFQRLMGIFTDGLRRSLASGYGSEVLSLEKVMTA
NPKCITEDSDIALALQIMRESSPVAVLPLVDNEENRHVTGLLHHHTLAKAGLL"
complement(7843. .9024)
/gene="sucB_2"
complement(7843. .9024)
/gene="sucB_2"
/codon_start=1
/transl_table=11
/product="dihydrolipoamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/transl_table="MIFEFPPKIGETSSGGSIVRWLKNLGDHVADEPLIEVSTDKI
ATLSPKAGRLVFCVQNEGDEVASDVLGLIELEISEADDETSCLPTSCETKSEA
SSSSSVSPAVLSAQREGIDGLNQLKTAGTGKGRVTRQDLEAYISQOQVSIPE
IFQGVNRIIPMSPLRAIASLSKSDVEPHASLVVDVDTDLNMLISGRQRFLOTH
GKVLITTSFTVQCLAQTLRQFPLNLGSDGTTVMKKSNNVGVAVNLNKGVVVPIH
NCQDRGLTSVALADALSSRARLNKLDSPSEVQGVSTVTFNFGMIGALIGMPIIRIPEV
AIIIGIGITQIRVVYRDDDSLAIKRMVYVTLTFDHRVLDGIYSEFITSUKNRLESVTM
G"
complement(9038. .10282)
/gene="glcT"
complement(9038. .10282)
/gene="glcT"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/transl_table="MKLWKFIFGLFVGTGLVLEDKAIFFPKIGDIFLNLLSMVY
PLVFCSMVLGIASIDMKKGRIGKSVGLYLGTTALTAIVIGLCFAWIFSPGNCDFPA
QAQSMDSAVTVIDSNKTAAYFLSIIAQVFPSPNVRSAFEGNIIQIITFAIFLGLRL
SGERGRPERFIDGFEIMLRVMYNNIMSFPYGVGASMAWISGNHGLVLMQGLKFI
AYTLACLPHATLVFGLVRFCKMSFSKFLSSNMDAISCASVASTASATLPVTRCYS
KNLGSAEVSQGVLPGLGATVNMNGTAIFOGMAAVFTAOAYNCPLSLSSLLLVVYATF
SAGVSAGVPGGMITLGSVLASVGLPIQGTAILAGIDRLDRIVGTPMNIIIGDAVATY
VASGEGELSPYESIKQSEVET"
complement(10282. .11379)
/gene="ycaH"
complement(10282. .11379)
/gene="ycaH"
/codon_start=1
/transl_table=11
/product="AtPase"
/protein_id="BAA98735.1"
/db_xref="GI:8978900"
/transl_table="MKKRPFTLFLFYRRVTIATLSLEGLGWGLGSLLSKVFALVLA
CMNRFSTPTYPARSTVTSVGNIVVGAGKPTVIMLAEALRLRGVSCGVLRSYKQSQ
SSRQKLTIVDSKVHSAASYVGDEPLMAEKLPGSVWVHKDRISARAAREKFGILL
DDGLQYKRLHKDVEIAVNGQDPLGGRAPFKRLRDFLRLLKTVDAITVNGGKEAG
TVKVRNSAPQIQRFTIASVWVTHGERIPKRALRELVRGVFCGLGPGQFLNMLRE
EGIHILGKYLIPDHAATTKKELAYFCQOMARQOGLLCTEKDVKSLPRLSGEVSLLP
IAKVMELSYNQDDTLSSLNMIQIHNKNGN"

```

Query Match 99.5%; Score 647.8; DB 1; Length 300550;  
 Best Local Similarity 99.7%; Pred. No. 6.4e-155;



[illegible]

	Qy	361	acagtcaggattatgtagtgcgaagccttgctcagttccagaataagatgttttcagggatt	420
	Dz	361	ACAGTACGTTATGAGTAGTGCGAAGCCCTTCGGGTCCAGAATAAGATTGTTTCAGGGATT	420
	Qy	421	ggccgctggtaattattaaagtgttgctgcccaagcaattgctgtctaactatgatcct	480
	Dz	421	GGCCGTGGTAATTATTAAAGTTTTGGTTCGCCCAAGCAATTGCTGCTAACATCATCCT	480
	Qy	481	aaaagcgctaagtgtttacaattataaaaggattttccgctctatatatgatggcatc	540
	Dz	481	AAAAGGCGTAATGTTTTACAAATATATAAGAGATTTCCGCCTCTATATGATGTCGATC	540
	Qy	*541	acagattctctatcatcacagaccttatgggcttactccaaccagcaaacgataaacctc	600
	Dz	541	ACAATTTCTCTATCATTTCCAGAGCTATTGGGCTTACTCCAACCAGCAAACGATAAATC	600
	Qy	601	ggcagtgattttactttccgaaagttagtctaggtagataaatctcagc	647
	Dz	601	GGCAGTGATTTTACTTTCCGAAGTTTGATCTAGGTATAATTTCAGC	647
	RESULT	6		
	AE001333/c			
	LOCUS	AE001333	10954 bp	DNA BCT 30-OCT-2000
	DEFINITION	Chlamydia trachomatis section 60 of 87 of the complete genome.		
	ACCESSION	AE001333	AE001273	
	VERSION	AE001333.1	GI:3329068	



KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

Chlamydia trachomatis.  
Chlamydia trachomatis  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
1 (bases 1 to 10954)  
Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,  
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
Koonin, E.V. and Davis, R.W.  
Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis  
Science 282 (5389), 754-759 (1998)  
9900809  
9784136  
2 (bases 1 to 10954)  
Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hymann, R.W.,  
Olinger, L., Grinwood, J., Davis, R.W. and Stephens, R.S.  
Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
Nat. Genet. 21 (4), 385-389 (1999)  
9920606  
10192388  
3 (bases 1 to 10954)  
Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,  
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
Koonin, E.V. and Davis, R.W.  
Direct Submission  
Submitted (20-MAY-1998) Program in Infectious Diseases, University  
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA  
Location/Qualifiers  
1..10954  
/organism="Chlamydia trachomatis"  
/strain="D/UW-3/CX"  
/db\_xref="taxon:813"  
complement(146..2089)  
/gene="CT622"  
complement(146..2089)  
/gene="CT622"  
/codon\_start=1  
/transl\_table=11  
/product="CHLPN 76kDa Homolog"  
/protein\_id="AAC68226.1"  
/db\_xref="GI:3329069"  
/translation="MESGFSVSSNQSMNPINQIASNETKESKESPASASS  
SVSSFLSSAKHALISLRDAIINKNSPTDSLSQLEASTSTVTRVAARDYNEAKS  
NYDTAKSLLENATLAEYETKMDALQMERLAKQAEVTRIKALQKQVDRK  
LNVLKQKQNTLKTFTTDSADQIPAINSOLEINKNSADQIKLEGNQISYEAV  
LTNAEVIKASEAGIKLQALQSIVDAGDSQAQLQAQONNSPDNIAATPKLIDAA  
ETKNELQKHTGLDPLVKKAEQISOAQDQIEIKPSGSDIPVPGSSAASAGS  
AVGALKSSNSGRISLDDVDNEMAAIAMQFPMIEQFNWNPATAKELQAMEAQL  
TAMSDQLYGDELPAETQAIKDALAQKQPSDGLATAMQVAFAAKVGGSAGT  
AGTVOMYKLYKTAFTSSSSAAALSDGSAYKTLNLSYSESRGVSQSAISQTAN  
PALSRVSRSQSGRSADASQRAAETIVRDSOTLGDVYSRLQVLSLMSITVSNPO  
VNOEIMOKLTASISKAPQFCYPAQNSADSLQFAAQLEFEFVDGERSLAESRENAF  
RKQPAFIOQVLNIASLFSGLS"  
complement(2197..3495)  
/gene="CT623"  
complement(2197..3495)  
/gene="CT623"  
/codon\_start=1  
/transl\_table=11  
/product="CHLPN 76kDa Homolog"  
/protein\_id="AAC68227.2"  
/db\_xref="GI:6578109"  
/translation="MKFYFYKFGVGLLACGSLNLAFAQASSMSQSLSVEDLDSYL  
SSKGFVETKRGDVLRLAGDVRARWYAKDELTTQTPAKPMLPTNRYSEFNLYVDY  
TANSWMTSKMNVIT IAGSSAAGLDINRAFLGYRYKNPETOAEVFAEIGRSLGD  
IFDSVQNSNEDGIIHLYAARRISSEKLPFTMIYHGGPFVNNMAKEYAWVEAILNKL  
PQNFVYKTSVDWNTLTAKNDPDASAAQPAKNTKYDILVWQLVGKSTAMPWFNG  
ATVRYEYLAIPEDVAGIGRGNQMYWFAQIKQGLDPKESNGFTNYKGVSTQFY  
MGLTDSVGFRAVAAYSKPDANLGSDEFTYRYKVDLGLISF"  
complement(3738..5348)  
/gene="mviN"  
/note="CT624"

complement(3738..5348)  
/gene="mviN"  
/codon\_start=1  
/transl\_table=11  
/product="Integral Membrane Protein"  
/protein\_id="AAC68228.1"  
/db\_xref="GI:3329071"  
/translation="MRKDDGSLVSLNLLSGTFFSRULTGMLREIVMATYFGADPLV  
ASFWLAFRIFFLRKLLGSPILGLAFIPHEFLRAQNISRAFFSFRRFCYSAIL  
FTLIIELGVCWCSCVSGISFDLLTLTIILFSGIFLMMYTNSTLLHCEKFFSVGL  
APSVNWSWIGTVFLARNYDPRNRIFGLAVLVIGFLEWAVTLPGVMKFGOSKEVP  
QERDSIRALIAPLSLGLSMGIFQLNLLCDMLARYINEVPLYLMSYRIQOLPVL  
FGLGVFTLLPAISRCVQDQHQOQYDILRFSLKTVAVVWMTMLLLFALPGVRVL  
YHGVFPKTAHVAIVEVLGSGSIIPMALAPVSAFVARRNYKVPMLVGIIAAVN  
MYNLGCLVCKQVAVLAYATSLVSGOLAMLVYCAKSLPTKGMWTEKESGKTIV  
ITILAAVITIGVNIHTHYVVFIEPLVPTKPLVSLFDQCGVFAESALFLSVLFG  
LAKLLTEDLMNLISFYWKGHOSILRN"  
5521..6387  
/gene="nfo"  
/note="CT625"  
5521..6387  
/gene="nfo"  
/codon\_start=1  
/transl\_table=11  
/product="Endonuclease IV"  
/protein\_id="AAC68229.1"  
/db\_xref="GI:3329072"  
/translation="MEILPPQEAALLGAHTSAAGLHNALEYGRDIGNATVOLFATNQ  
RQKRRITLQWVDQFRIALNETSLYSINSHAGYLNPGAPNPNLEKTRVCMHQEIA  
DCISLGSFVNFHFGAALSDEKSLDRTIASFSQMAPFENPNPVLVLETTAGGGS  
LIGSFEELAYLQIGKHIPICVCLDTCHIFAAGYDISVAGWESVVLKHFDAVIGLS  
FLRAIHLDVSFALGKNKRHAPIGEGICSGISDFCFLMQDERTMLPKYLETPGGPDL  
WTKIERYLQKVC"  
complement(6484..7113)  
/gene="rs4"  
/note="CT626"  
complement(6484..7113)  
/gene="rs4"  
/codon\_start=1  
/transl\_table=11  
/product="S4 Ribosomal Protein"  
/protein\_id="AAC68230.1"  
/db\_xref="GI:3329073"  
/translation="MARYCGPKNRIARRFGANIFGRNPLLRKPNPQGHQWRRKK  
SDYGLQLEERQKLCACYMILEKQKAYEVNKNQVNAQMELEFECRLDNIVYRL  
GFAKTIFAQQLVSHGHVLVNGKVRDRSFVRPQMOISLKEKSKRLAIYTESLENKD  
QSSLPVLSLDKAAFGELVVAPELDQIASQLPLPVNVSVICFELSHRT"  
7497..8480  
/gene="ycea"  
/note="CT627"  
7497..8480  
/gene="ycea"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC68231.1"  
/db\_xref="GI:3329074"  
/translation="MEKNYALAYVEYFVSNPHEEIALHKOLFKTMDVSCRIVISEE  
GINGQFSYQOPDAERYNAWLQRPDFAIKFHHIENIFPRVTYKRELVALGCS  
VDHTKQKHISPEWHEKLOENRCLVDVRNNYEMWIKGFENAVLPDIETFRFPDYA  
DRLAKEHDPKTPVMYCTGIRCELYSALLLEKGFKEYQDGGVLYGLKMGTKGW  
RGKILFVDRRMATIDEADPNVSPFIARCSLCTDSTYINCANTCNLCNLTCCESCLT  
THKCCSEESQAPRIAFSAERGNKPFRRKHLCTPTIEQSCCLKEQENQPA"  
complement(8552..9427)  
/gene="ispa"  
/note="CT628"  
complement(8552..9427)  
/gene="ispa"  
/codon\_start=1  
/transl\_table=11  
/product="Geranylgeranyl pyrophosphate synthase"  
/protein\_id="AAC68232.1"  
/db\_xref="GI:3329075"



Db	2472	GTTTGAAGCTTTACGCGATTCCAGAAATTGATGTCGGGTATTGTCGGGAACCAAAATG	2413
Qy	439	aegtttttggtctgcgcgaagcaaatgtcgtctaactatgatcctaaaggagctaaatgagtttt	498
Db	2412	AAATATTGGTTTGCTCAAGCTATCAACAAGGATTGGATCCTAAAGAAATCTAAACGGCTTT	2353
Qy	499	acaaatataaaggagtttccgcctctatatatgtatgcatacagagattcctctcatcatt	558
Db	2352	ACTAACTATAAAGGAGTTTCCTATCAGTTGTATTATGGGTCTGCACAGATTTCGTTTC	2293
Qy	559	agagcttatg99gcttactccaaacagcaaaacgataaactcggcagtgatttcacttcc	618
Db	2292	CGAGCTTATGCTGCTTTATTTCTAAAGCTGCTCAACGATAACCTTGGTAGGACATTCACCTAT	2233
Qy	619	cgaagcttgatctagggtataaattccagcgtt	650
Db	2232	CGTAAGTATGACCTAGGTTTAATTTCTTCATT	2201

AE002357/C	AE002357	14698 bp	DNA	BCT	26-MAY-2000
LOCUS					
DEFINITION	Chlamydia muridarum, section 84 of 85 of the complete genome.				
ACCESSION	AE002357	AE002160			
VERSION	AE002357.2	GI:8163344			

SOURCE	ORGANISM
Chlamydia muridarum.	Chlamydia muridarum.
Chlamydia muridarum	Chlamydia muridarum
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

ORGANISM

Chlamydia chlamydiae  
Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.  
1 (bases 1 to 14598)  
Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F.,  
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,  
Berry, K., Backs, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J.,  
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.

REFERENCE

AUTHORS

**TITLE.** Genome sequences of *Chlamydia trachomatis* MoPn and Chlamydia pneumoniae AR39

**JOURNAL** Nucleic Acids Res. 28 (6), 1397-1406 (2000)

20150255  
MEDLINE  
PUBMED  
REFERENCE  
2 (bases  
1 to 14698)  
Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F.,  
White, O., Hacks, E.K., Peterson, J., Umayam, L.A., Utterback, B.,  
Berry, K., Bick, S., Linher, K., Weidman, J., Khouri, H., Craven, T.,  
Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R.,  
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.,  
Kolony, J.,

White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weldman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.

TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
COMMENT On Jun 1, 2000 this sequence version replaced nt-710q033

```

FEATURES
  source
    1. .14698
      /organism="Chlamydia muridarum"
      /db_xref="taxon:83560"

```

```

gene
/notes="synonym: Chlamydia trachomatis MOpn"
/nc_xref="taxon:83300
complement(105. .1364)
/gene="TC0908"
complement(105. .1364)

```

complement(103. .1304)  
/gene="TC0906"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/aa=224

```

/cocon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39699.1"

```

```
/db_xref="GI:7190934"
/translation="MWGFLKQIGFRERSRVGKQFYHCVSRYFHYLPVLAILL
SWPFLSEQQWRVSYFLFPVSSJGWLFAIGLRERQLRAAGOLLETIRKLTED
KNTRTETIRKQKQANRLK LUNDKLVRLGQAREVYIQAKGRYDHMESEKSKLHEE
```

LQWLEAAVRENEKTLLENQELRQELKETLAYOQELHDEYQATFVEQHNLMDRRQ  
GNLEAKVQDLMCSRLNLLQVEGAKTELPGRMSRDVVAQLVLEFKRIYFRVET  
ADSLTALRYTRTPDSAHNYSLACRQLFDDLEENLGMFIYAPQARVLFANALF











```

Quality coverage: 4-7 in Q20 bases; agarose-fp
Quality coverage: 4-8 in Q20 base.
NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
      I
      * 2771: contig of 2771 bp in length
      *
      * 2772 2871: gap of 100 bp
      * 2872 6938: contig of 4067 bp in length
      * 6939 7038: gap of 100 bp
      * 7039 15741: contig of 8703 bp in length
      * 15742 15841: gap of 100 bp
      * 15842 38593: contig of 22752 bp in length
      * 38594 38693: gap of 100 bp
      * 38694 60937: contig of 22244 bp in length
      * 60938 61037: gap of 100 bp
      * 61038 93236: contig of 32199 bp in length
      * 93237 93336: gap of 100 bp
      * 93337 143364: contig of 50028 bp in length.
FEATURES             location/Qualifiers
source               1..143364
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-45J21"
                     /clone_lib="RPCI-11 Human Male BAC"
misc_feature         1..2771
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
misc_feature         2872..6938
                     /note="assembly_fragment"
misc_feature         7039..15741
                     /note="assembly_fragment"
misc_feature         15842..38593
                     /note="assembly_fragment
                     clone_end:SP6"
misc_feature         38694..60937
                     /note="assembly_fragment"
misc_feature         61038..93236
                     /note="assembly_fragment"
misc_feature         93337..143364
                     /note="assembly_fragment"
BASE COUNT   45788 a 25396 c 25729 g 45848 t    603 others
ORIGIN

Query Match          7.1%; Score 46.2; DB 2; Length 143364;
Best Local Similarity 51.7%; Pred.No. 0.18;
Matches '105; Conservative 0; Mismatches 98; Indels 0;

QY      64 ttgtgaaatgtagtggctgcgaactggacacattcgttcccttcagaacctccaactac
Db      34837 TGTGTATAAATTAGACTACTTAGATAAAATATATGCAATTTTGGAGAACCTCAATAAC
QY      124 gaaaagctgctacaaacgctatgaaatacacaataactgtgttgaggcgctgcgggg
Db      34897 AAACCTGATTCAAGAAAAAATAGAAAATACTCATAGATTTCTPACAAGTAAGAATATTCG
QY      184 agcatagtcaggttcccttgatcaatgacagaaaaaacctctatatcttatgagac
Db      34957 ATTAGTAATAAAGTAGTACTTTCCACAAAGAAAAACATCAGGCCCATATTTCTTTATTCGAG
QY      244 ttcttaatgaaccctttagcaaa 266
Db      35017 ATTCTATCAAATACTTAAGTAAA 35039

RESULT 10
```



```
AX083744/c
LOCUS       AX083744             1141 bp    DNA             PAT      28-FEB-2001
DEFINITION   Sequence 22 from Patent WO0111061.
ACCESSION   AX083744
VERSION      AX083744.1   GI:13185472
KEYWORDS
SOURCE       synthetic construct.
             synthetic construct
             artificial sequence.
REFERENCE    1 (bases 1 to 1141)
AUTHORS      Kunst,R. and Clemens,S.
TITLE        Regulation of embryonic transcription in plants
JOURNAL      Patent: WO 011061-A 22 15-FEB-2001;
             UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES
             Location/Qualifiers
             source
               1..1141
               /organism="synthetic construct"
               /db_xref="taxon:32630"
               1..1141
             promoter
               /note="consensus sequence of A.t., L.a., and B.n. PAEL
               promoters"
BASE COUNT   123 a    32 c    42 g    112 t    832 others
ORIGIN
Query Match      6.9%; Score 44.8; DB 6; Length 1141;
Best Local Similarity 10.6%; Pred. No. 0.37;
Matches 48; Conservative 189; Mismatches 215; Indels 0; Gaps 0;
QY 55 aaacagcttttggtaaatgtagttgtcgactggaacacattcgtcttcacgaacc 114
DB 519 HNNNTDTRYWWKRWABRTTYVDSMCAKSMWRGNWRAMKMWMAANNNDAGMDHW 460
QY 115 tcactacagaaagctgctacaacgcctatgaatacaatactgtgttgcgagtgg 174
DB 459 TWMGNNTWWRAMKMMWMAKRCRAYCCNNNNNACVWHKHKWRTWTKYMKAAACNNN 400
QY 175 ctctcgaaacagcatagtcaggtctcttgatcaatgacagaaagcctctatatctt 234
DB 399 NBKAMYMRVAMMYSRDNTTDMWMTSDWBWHVTVDTYTMKRWNNNNNNNRBCKTTS 340
QY 235 tatgagctttcttaataaacctttagcaaggtctagcaagactacgttaaatgaaaa 294
DB 339 WWWMDHMHNTCTYGNNTWGSAYBAWMAWSMAAGASBNBYTWCMRMTYMGKTMNTNNNN 280
QY 295 gaaacactagctgtgttattgaggagacatttaggggagactcagaaagctggagactgg 354
DB 279 NKAWYRTKTVACNNRYYYDTAVTBRNKYCYAYBYWYBMWCKHHWWRABHRS 220
QY 355 tctgcacagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 414
DB 219 WNMWVVKCNKYMVSWHYHAMRYBKWABAVGCGNNWKKDRMAHHWCATNNNNMMWYAYM 160
QY 415 gggattggcgcgtgtaattatttaagtttgcgttcgcgaagcaatgctgctgaactat 474
DB 159 HHMKKGKRAWTNNKTABRDDHBAHVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 100
QY 475 gatcctaagagcgttaagtcagttttacaaatta 506
DB 99 DRYVSANNTGVRWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 68

RESULT 11
AP001118
LOCUS       AP001118             347550 bp    DNA             BCT      07-SEP-2000
DEFINITION   Buchnera sp. APS genomic DNA, complete sequence, segment 1/2.
ACCESSION   AP001118 AP000398 BA000003
VERSION      AP001118.1   GI:10038693
KEYWORDS     ferredoxin II; GRP-binding protein; membrane protein; kinase.
SOURCE       Buchnera sp. APS (strain:APS, isolate:Tokyo1998,
             specific host:Acyrthosiphon pisum) DNA.
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Buchnera.

1 (sites)
Shigenobu,S., Watanabe,H., Hattori,M., Sakaki,Y. and Ishikawa,H.
Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS
Nature 407 (6800), 81-86 (2000)
20445173
2 (bases 1 to 347550)
Shigenobu,S., Hattori,M., Watanabe,H., Toyoda,A., Yada,T.,
Sakaki,Y. and Ishikawa,H.
Direct Submission
Submitted (24-JAN-2000) to the DDBJ/EMBL/GenBank databases. Shuji
Shigenobu, University of Tokyo, Department of Biological Sciences;
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
(E-mail:shige@gs.riken.go.jp, URL:http://buchnera.gsc.riken.go.jp,
Tel:81-3-5800-3553, Fax:81-3-5800-3553)
Collaboration Information:
RIKEN, Genomic Sciences Center (GSC)
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan
Tel:81-42-778-9923 FAX:81-42-778-9924
AP000398: Submitted(23-Aug-1999).
Location/Qualifiers
1..347550
/organism="Buchnera sp. APS"
/isolate="Tokyo1998"
/species="Buchnera sp."
/strain="APS"
/db_xref="taxon:107806"
/note="primary endosymbiont of Acyrthosiphon pisum"
197..2083
/gene="gida"
197..2083
/gene="BU001"
/codon_start=1
/product="glucose inhibited division protein A"
/protein_id="BABI2729.1"
/db_xref="GI:10038694"
/translation="MFNLNFDVIVVGAGHAGTAAASSBMCKTLLLTOKISDLGA
LSCPAIGGICKSHLYKEIDALGMMKADIDYSGIOFRILNNSKGPVRSRAQADKI
LYHETVRKILKKQNNLLIEAEVKDLIFKNYSVVGVLQNEINFYRSVSLAAGTFLG
KIHGLKYSAGRIGDKSAIDLSVRLRELRLNRKLTGTTPRIDINTVFNFNLLIQ
NSDPVPVSEFMGNVSHHPKIQPCYLTHTNEKTHEIRKNDKSPITGFLKGLGPRY
CPSEDKIVRFPDRKSHQVLEPEGLSSIKVYPNGISTSLPEVEQOIVASTKGLKES
KIIIRPVAIEYDFDPKDLNLTLESLKIKLGFAGQINGTTGYEPAASQGLLAGLNA
LSSKNTGEMFRDQAVYGLVDLDTTQCTEPEYEMFTSRAEYRLSLREDNADRLTE
IGRLGLVNDSEWRYNOKVLIQTEMNRLKKNKISPIDADILKKLYNINLKEIS
MSELKPOIRYQDLQSLSEFTGLVDLEAGQIENEIKYAGYIKRQSEETERHLKNE
NTFLSSYDYNKIRGLSSEVYKLNLDYKPISIGQASRISGITPAISILLHLKESY
KHTL"
2278..3102
/gene="atpB"
2278..3102
/gene="atpB"
/EC_number="3.6.1.34"
/note="BU002"
/codon_start=1
/product="ATP synthase A chain"
/protein_id="BABI2730.1"
/db_xref="GI:10038695"
/translation="MILEKISDPQKYISHHLSHLQIDLRSPKTIQPGALSSDYTVNV
DSMFFSLVGSFFELISFYMWGKIKTQGPGLQTAIELIFFEVLNVKSMYQGNKALI
APLSLTVFWFLNMLMDLVDFDFFSEKVFELPAMRIVPSADINTITLSMGLGVFF
LILFTVKIKGYVGLKELILQPNHPVSEIFNFIIEVSLVSKPIISGLRLFGNNYA
GEMFILLAGLLPWTQCFNLVNPWNAIFHILISLQAFIMVLTIVYLSMASOSHKD"
3139..3378
/gene="atpE"
3139..3378
/gene="atpE"
/EC_number="3.6.1.34"
/note="BU003"
```



```

/codon_start=1
/transl_table=11
/product="Atp synthase C chain"
/protein_id="BAB12731.1"
/db_xref="GI:10038696"
/translation="MENLNDMLYIAVAIMVGLASIGTAATIGLGGKFLGGAARQPD
LVPLLRTQFFVVMGLVDALPMIAVGLGYMLFAIS"
3497..3982
/genes="atpF"
3497..3982
/genes="atpF"
/EC_number="3.6.1.34"
/note="BU004"
/codon_start=1
/transl_table=11
/product="Atp synthase B chain"
/protein_id="BAB12732.1"
/db_xref="GI:10038697"
/translation="MNLNATIIGQASIFVLFWFCMKYIWPPIIALIETROKIKESL
TNAKARDELYILEKIHQNIIDAKQASNLINSAKOKVSILEDARNQALESKKLI
LNTQSEINIAITHARKNLHKEVDLSISMAEKLIITKNISKDDNOELDELVTLSQVK
N"

```

LFNRYTESOVQSIILENTASEHAARMAMTKTATNSGNRIKELQLVYNKVRQANLIQ  
LNEIVSAGAVSID"  
6997 . 8394  
/gene="atpD"  
/gene="atpD"  
6997 . 8394  
/gene="atpD"  
/EC\_number="3.6.1.34"  
/note="B0008"  
/codon\_start=1  
/transl\_table=11  
/product="ATP synthase beta chain"  
/protein\_id="BAB12736.1"  
/db\_xref="GI:10038701"  
/translation="MATGKIIQIIIGAVDVFNQDSVPKIIYALNVEGNKOYKLIIEVQ  
QOLGSGVVRTIAMGSSNGLKRLGIIVTDLGHYIKVPGVEATLRGLINVLGETINDKAL  
KNQSDIKIYEWHITRSPPTSRQOASCRELETGKIIVDILCPFSKCKVSLVYCGAVG  
KTNVMELIENIAHVSQVSGTGVGRTRENGTFHMDNSQVDKVKSLVYCGMNEP  
PGNRLVAFGLIAFKRFEQDKVLLFDINIRYTLHAGTSEVALLGRMPSAVQIOT  
KAMEGLLQNSTIKTSKNGITSQVAYVPADLTPSPATTFHLDSTVLSLQIASL  
GIYPAIDPLASTSRQLDPKYIVGDEHYETAVQLQSTLYRQELKDI IAILGMDLASEL  
KLVSRARKIQRLSQSPFFVAEVTGFCPKYVYSLKDNIRAFKGIKEGFDDLPQPAQY  
NNGSVIEVTEKAKKI."

```

Nn
3982. .4515
/ gene="atpH"
3982. 4515
/ gene="atpH"
/ EC_number="3.6.1.34"
/ note="BU005"
/ codon_start=1
/ transl_table=11
/ product="ATP synthase delta chain"
/ protein_id="BA812733.1"
/ db_xref="GI:10038698"
/ translation="MSVADT IARYPQAQAFETAIENNTIEKWNLLFIKTIASHKFF
KNFSGISPKYLSLFIPTIGTNIIDENAKNIKLISNQRFNLNITFERFVKLEAC
YKNTIIIVQLSAFSLKLINKRKVLERFFLKKTKIYKYVDPNLLINGMIYKVNTTIF
DESRNHLKOLDSINP"

```

```

8421.. 8837
/ gene= "atpc"
8421.. 8837
/ gene= "atpc"
/ EC_number= "3.6.1.34"
/ note= "HU009"
/ codon_start= 1
/ transl_table= 11
/ product= "NTP synthase epsilon chain"
/ protein_id= "BAB12737.1"
/ db_xref= "GI: 10038702"
/ translation= "MNFYLDVSLTKTIFSGVEKIRVSGSEGELGIYPGHAQLLSIL
RPGWYIYFHKRKEECIYISGGILEVQSVSVVIAHAIDLRSDIRIKTKTAKNEE
ESIKGNNTKINKDAILLQISKEAKLRVLEVMDFK"
complement(8911.. 11322)
/ gene= "gyrB"
complement(8911.. 11322)
/ gene= "gyrB"
/ EC_number= "5.99.1.3"
/ note= "HU010"
/ codon_start= 1
/ transl_table= 11
/ product= "DNA gyrase subunit B"

```

```

4530..5068
/ gene = "atpA"
4530..5068
/ gene = "atpA"
/ EC_number = "3.6.1.34"
/ note = "BU006"
/ codon_start = 1
/ transl_table = 11
/ product = "ATP synthase alpha chain"
/ protein_id = "BBA12734.1"
/ db_xref = "GI:10038699"
/ translation = "MKLNSTETSKLKKIRIAQFEVFNQSYNQSNGVTSIVSDGIITKINGL
SWLVGMGLTPNNFYAIALNEDTDFVAVGVMPYTHISEGAKVCTGKILFVPGDN
FLGRVNALGPIIDGCKTQDFVPEADPGVIDHKSVPNOIQGYIVADSMIPDG
RQREILLIGDRGTGKTALAIDTIIINQKSGIKCIYVACGKLUKTIINVKVLENNAL
FNTIIVVASASAAQLYLAIPYSGCAMAFPRFKNGDSIIIVDLSKHAVAYKISLL
LRPPGAPFPGDIFVLSKRLLEASRVSMVEYVQKTKNKTKITGSGTALPIETQFS
GVDSVAPVNTPIITDQIPLFNLNFSGIRPAVNPGISVSRVGSAAQTIITKLSGG
ITRALAYVELAAAFSSFDLDDTTKQLNHGKQITELLKQKQSPISIAEQALILFY
AANNELPOLYSEKIPKEPEKELIYAHVYDPEFETKNGCDENIVIKDKAFIKLITFEF

```

[illegible]

```

KNOF"
6101..6973
/genes="atpG"
6101..6973
/genes="atpG"
/EC_number="3.6.1.34"
/note="BU007"
/codon_start=1
/transl_table=11
/product="ATP synthase gamma chain"
/protein_id="BAB12735.1"
/db_xref="GI:10038700"
/translation="MSTREIKNKIVSTVTKITKAMENVAVSKMRKTEEMRSRGP
YSDIIRKVIDHVQGNLEYKHSYLEERKTNRIGMIIISTDRGLCGGLTNTLFRQVLPK
IQNFAVTPCDIILRLGLKSLVFKLGCSNLAKNGLNPGNKEELINSVGLIQEY
OAKRKDIFTAYAFNKFMKMSOYPTITQLLPFSKKNDQASNNNDWYIEEPSKLIIDY

```

[illegible]



<b>KEYWORDS</b>					
SOURCE					
<b>ORGANISM</b>					
Human.					
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
<b>REFERENCE</b>					
<b>AUTHORS</b>					
Waterston,R.H.					
<b>TITLE</b>					
The sequence of Homo sapiens clone					
<b>JOURNAL</b>					
Unpublished					
<b>REFERENCE</b>					
<b>AUTHORS</b>					
Waterston,R.H.					
<b>TITLE</b>					
Direct Submission					
<b>JOURNAL</b>					
Submitted (16-MAR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA					
<b>REFERENCE</b>					
<b>AUTHORS</b>					
Waterston,R.H.					
<b>TITLE</b>					
Direct Submission					
<b>JOURNAL</b>					
Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA					
<b>REFERENCE</b>					
<b>AUTHORS</b>					
Waterston,R.H.					
<b>TITLE</b>					
Direct Submission					
<b>JOURNAL</b>					
Submitted (25-MAR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA					
<b>COMMENT</b>					
On May 29, 1999 this sequence version replaced gi:4572701.					
<b>FEATURES</b>					
source					
1..174327					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone="RP11-71F18"					
BASE COUNT 57722 a 31892 c 30845 g 53868 t					
ORIGIN					
Query Match 6.2%; Score 40.4; DB 9; Length 174327;					
Best Local Similarity 51.1%; Pred. No. 5.6;					
Matches 95; Conservative 0; Mismatches 91; Indels 0; Gaps 0;					
QY 390 gtccagttcagaaaatagatttcaggagtgccggtggtaattttaaagtgttgggtt 449					
Db 46387 GCCAGATACGACTAAGTGGATTGGTTGGTTGTATAGATCATCTGAATAATTAAGT 46446					
QY 450 cgcccagaacaaatgctgcgaactgatcctaagaagcgctaatagttttacaattataa 509					
Db 46447 CCTCAATATCATGCCCTTGTTTAAATACATGAAGAGTGATGAAGTCACCACACTCAGT 46506					
QY 510 aggattttccgctctatatatgatggcatcacagattctcttatcattcagagttatgg 569					
Db 46507 TGCATTTTGGTATTATCTATTCTGTAAATTTCCACACTTTGTAAATTAATGAGATAG 46566					
QY 570 ggctta 575					
Db 46567 AATTTA 46572					
RESULT 13					
AL451103/c					
LOCUS					
DEFINITION					
Human DNA sequence from clone RP11-172M3 on chromosome Xp21.1-21.2,					
complete sequence.					
ACCESSION					
AL451103 AC024521					
VERSION					
AL451103.7 GI:13751498					
HTG.					
KEYWORDS					
human.					



\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.



misc\_feature /note="assembly\_fragment"  
104407..122860  
misc\_feature /note="assembly\_fragment"  
122961..149152  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:left  
149253..179033  
misc\_feature /note="assembly\_fragment"  
BASE COUNT 56286 a 32504 c 32567 g 56075 t 1601 others  
ORIGIN

Query Match 6.0%; Score 38.8; DB 2; Length 179033;  
Best Local Similarity 47.5%; Pred. No. 14;  
Matches 115; Conservative 0; Mismatches 127; Indels 0; Gaps 0;  
QY 310 ttatttgagggaacttttgggggaactcagaaaagctgagactgtctgcccacagtaagt 369  
DB 135097 TTTTCTGATGAATATATATCGCTATGCAGTGTATCAGTAATATGTTTGCCTTAAAGTCT 135038  
QY 370 tatgagtgtcgaagcccttgcttcagaaatagatgtttcagggaattggccgtggt 429  
DB 135037 GTTCTCTGTAGCTACTACTACCTTCTTGTAGTCAGTATTTTCACAGAGTAACCTCTTT 134978  
QY 430 aattattaaagtgttggttcgcccaagaactgtgcttaactatgactcctaaagaggt 489  
DB 134977 CATCTATTTAAATTTAAATATTTCTGTATCTTTTAAATGTATCTGCTAAACAA 134918  
QY 490 aatagttttacaaataaagaatttcgcctctatatgtatggtcattcagattct 549  
DB 134917 AATAGATTAGGATTTTGTAGTGAGTGTGTTTGTGTTTGTATTTTGTAGTCCAGAGGTCT 134858  
QY 550 ct 551  
DB 134857 CT 134856

RESULT 15  
HS70501/c  
LOCUS HS70501 161888 bp DNA PRI 18-MAR-2001  
DEFINITION Human DNA sequence from clone RP4-70501 on chromosome 20q11.2  
Contains STSS and GSSs, complete sequence.  
ACCESSION AL035251  
VERSION AL035251.11 GI:4585773  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Laird.G.  
Direct Submission  
Submitted (09-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Apr 16, 1999 this sequence version replaced gi:4584645.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP

database can be found at  
http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr20  
RP4-70501 is from the library RPCI-4 constructed by the group of  
Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RP4-70501.

FEATURES  
Location/Qualifiers  
1..161888  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="q11.2"  
/clone="RP4-70501"  
/clone\_lib="RPCI-4"  
636..836  
/note="L2 repeat: matches 2349..2567 of consensus"  
1919..2158  
/note="5 copies 48 mer 71% conserved"  
1924..2145  
/note="74 copies 3 mer tga 73% conserved"  
2220..2671  
/note="MLN1C repeat: matches 6..466 of consensus"  
3649..3824  
/note="8 copies 22 mer 84% conserved"  
4351..4763  
/note="L2 repeat: matches 2292..2734 of consensus"  
5118..5195  
/note="MLN1J repeat: matches 320..392 of consensus"  
5491..5603  
/note="MERSB repeat: matches 64..176 of consensus"  
6011..6070  
/note="15 copies 4 mer gaga 71% conserved"  
6014..6075  
/note="31 copies 2 mer ag 72% conserved"  
7369..7433  
/note="MIR repeat: matches 48..114 of consensus"  
8095..8290  
/note="4 copies 49 mer 88% conserved"  
8216..8311  
/note="2 copies 48 mer 95% conserved"  
8610..8909  
/note="150 copies 2 mer ct 55% conserved"  
8637..8959  
/note="19 copies 17 mer 58% conserved"  
8649..9032  
/note="8 copies 48 mer 61% conserved"  
8689..9031  
/note="7 copies 49 mer 68% conserved"  
8724..9011  
/note="72 copies 4 mer cctt 60% conserved"  
8842..8943  
/note="34 copies 3 mer ctc 62% conserved"  
9435..9491  
/note="MIR repeat: matches 83..145 of consensus"  
9560..9751  
/note="MIR repeat: matches 11..211 of consensus"  
9959..10286  
/note="164 copies 2 mer gg 62% conserved"  
9978..10222  
/note="5 copies 49 mer 71% conserved"  
9979..10266  
/note="6 copies 48 mer 68% conserved"  
9986..10291  
/note="18 copies 17 mer 63% conserved"  
10041..10282  
/note="11 copies 22 mer 66% conserved"  
11402..11452  
/note="3 copies 17 mer 90% conserved"



repeat_region	12586..12667 /note="LIMB8 repeat: matches 6194..6290 of consensus LIMB8 repeat: matches 6194..6290 of consensus"
repeat_region	13217..13513 /note="ALUSg repeat: matches 1..294 of consensus"
misc_feature	14979..15012 /note="match: GSS: Em:AQ424325"
repeat_region	15211..15965 /note="LIMB3 repeat: matches 5447..6182 of consensus"
repeat_region	15995..16440 /note="MLT1H repeat: matches 29..547 of consensus"
misc_feature	16441..16586 /note="match: GSS: Em:AQ072870"
repeat_region	17144..17261 /note="MLT1C repeat: matches 428..534 of consensus"
repeat_region	18488..18530 /note="Charlie6 repeat: matches 1128..1171 of consensus Charlie6 repeat: matches 1128..1171 of consensus"
repeat_region	18554..18760 /note="MER58A repeat: matches 9..216 of consensus"
repeat_region	18795..19024 /note="L2 repeat: matches 2338..2581 of consensus"
repeat_region	19440..19473 /note="L1P repeat: matches 503..534 of consensus"
repeat_region	19862..19959 /note="L2 repeat: matches 2581..2702 of consensus"
repeat_region	20808..20855 /note="12 copies 4 mer acac 89% conserved"
misc_feature	21699..22143 /note="match: STS: Em:G09929"
repeat_region	21766..21916 /note="MER3A repeat: matches 9..189 of consensus"
repeat_region	23787..24263 /note="MLT1C repeat: matches 22..466 of consensus"
repeat_region	24290..24562 /note="LTR16C repeat: matches 117..385 of consensus"
repeat_region	24988..26483 /note="L1M1 repeat: matches -1390..105 of consensus"
repeat_region	28153..28259 /note="MIR repeat: matches 36..176 of consensus"
repeat_region	28986..29049 /note="16 copies 4 mer cact 79% conserved"
repeat_region	30039..30458 /note="L2 repeat: matches 2112..2519 of consensus"
repeat_region	30691..30931 /note="MIR repeat: matches 39..256 of consensus"
repeat_region	30934..31035 /note="34 copies 3 mer aat 68% conserved"
misc_feature	31036..31210 /note="match: STS: Em:G09813"
repeat_region	31447..31520 /note="L2 repeat: matches 2632..2705 of consensus"
repeat_region	32120..32216 /note="MLT1H repeat: matches 432..536 of consensus"
repeat_region	32346..32575 /note="MER20 repeat: matches 1..218 of consensus"
repeat_region	32620..32727 /note="27 copies 4 mer cctt 63% conserved"
repeat_region	32729..33075 /note="L2 repeat: matches 2294..2674 of consensus"
misc_feature	complement(33076..33152) /note="match: GSS: Em:AQ489056"
repeat_region	33352..33819 /note="ALUSg repeat: matches 13..300 of consensus"
repeat_region	33820..33911 /note="L2 repeat: matches 2213..2308 of consensus"
repeat_region	34001..34100 /note="MLT1B repeat: matches 290..390 of consensus"
repeat_region	34261..34326 /note="3 copies 22 mer 86% conserved"
repeat_region	34267..34326 /note="30 copies 2 mer tg 88% conserved"
repeat_region	34273..34328

Search completed: February 7, 2002, 19:27:24  
Job time: 13158 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:46:11 ; Search time 685.35 Seconds  
(without alignments)  
814.356 Million cell updates/sec

Title: US-09-391-606-2

Perfect score: 651

Sequence: 1 atgacaaaaaacattatgc.....tagtataatttcagcgttt 651

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

```

1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	647.8	99.5	1550	21 AAA28411	Chlamydia pneumoni
2	647.8	99.5	2238	21 AAD02066	C. pneumoniae 76 k
3	647.8	99.5	1230025	20 AAX91990	Nucleotide sequenc
4	87	13.4	936	22 AAF58252	Oligonucleotide D1
5	87	13.4	936	22 AAF58254	Oligonucleotide D1
6	87	13.4	936	22 AAF58257	Oligonucleotide D1
7	87	13.4	936	22 AAF58259	Oligonucleotide D2
8	87	13.4	936	22 AAF58262	Oligonucleotide D2
9	87	13.4	938	22 AAF58255	Oligonucleotide D1
10	84.6	13.0	936	22 AAF58252	Oligonucleotide D1
11	84.6	13.0	936	22 AAF58254	Oligonucleotide D1

12	84.6	13.0	936	22 AAF58257	Oligonucleotide D1
13	84.6	13.0	936	22 AAF58259	Oligonucleotide D2
14	84.6	13.0	936	22 AAF58262	Oligonucleotide D2
15	84.6	13.0	938	22 AAF58255	Oligonucleotide D1
16	40	6.1	244	22 AAF58238	Oligonucleotide D1
17	38.2	5.9	15218	19 AAF18275	RSV isolate 2B w11
18	38.2	5.9	15218	20 AAZ22909	Nucleotide sequenc
19	38.2	5.9	15218	20 AAZ22914	Nucleotide sequenc
20	38.2	5.9	15218	20 AAX35267	DNA encoding the L
21	38.2	5.9	15219	19 AAV18277	RSV vaccine 2B33F
22	38.2	5.9	15219	19 AAV18278	RSV vaccine 2B20L
23	38.2	5.9	15219	19 AAV18279	RSV revertant 2B33
24	38.2	5.9	15219	19 AAV18280	RSV revertant 2B20
25	38.2	5.9	15219	20 AAZ22911	Nucleotide sequenc
26	38.2	5.9	15219	20 AAZ22912	Nucleotide sequenc
27	38.2	5.9	15219	20 AAZ22913	DNA encoding the L
28	38.2	5.9	15219	20 AAX35269	DNA encoding the L
29	38.2	5.9	15219	20 AAX35270	DNA encoding the L
30	38.2	5.9	15219	20 AAX35271	DNA encoding the L
31	38.2	5.9	15219	20 AAX35272	Oligonucleotide D1
32	36.8	5.7	244	22 AAF58238	Respiratory syncyt
33	36.6	5.6	15225	19 AAV17552	Respiratory syncyt
34	36.6	5.6	15225	21 AAA88744	RSV isolate 18537
35	36.6	5.6	15229	19 AAV18276	Nucleotide sequenc
36	36.6	5.6	15229	20 AAZ22910	DNA encoding the L
37	36.6	5.6	15229	20 AAX35268	DNA encoding the L
38	35.6	5.5	2334	18 AAT66428	Canarypox virus C5
39	35.6	5.5	2949	13 AAQ25764	Canarypox virus C5
40	35.6	5.5	3208	16 AAT04698	PRW/64.5 (contg. C
41	35.6	5.5	3208	17 AAT34124	Canarypox virus Pv
42	35.6	5.5	3208	18 AAT69990	Canarypox virus DN
43	35.6	5.5	3208	20 AAX79930	C5 ORF from canary
44	35.6	5.5	3209	13 AAQ22769	Canarypox PvUII fr
45	35.6	5.5	3209	13 AAQ29107	Canarypox PvUII fr

#### ALIGNMENTS

```

RESULT 1
ID AAA28411 standard; DNA; 1550 BP.
XX
AC AAA28411;
XX
DT 29-AUG-2000 (first entry)
XX
DE Chlamydia pneumoniae lorf2 coding sequence.
XX
KW lorf2; vaccine; antibacterial; antigen; ss.
XX
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 101..1369
FT /tag= a
XX
PN WO200024901-A1.
XX
PD 04-MAY-2000.
XX
PF 28-OCT-1999; 99WO-GB03565.
XX
PR 28-OCT-1998; 98US-0106037.
XX
PR 20-SEP-1999; 99US-0154658.
XX
PR 26-OCT-1999; 99US-0427501.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Murdin AD, Oomen RP, Dunn PL;
XX
DR WPI; 2000-350742/30.
DR P-PSDB; AAY92716.

```



XX Isolated polynucleotide encoding a Chlamydia polypeptide useful to  
PT treat, diagnose and prevent disease caused by Chlamydia infection  
XX  
XX  
PS Claim 1: Fig 1A-C; 88pp; English.  
XX  
XX This sequence encodes lorf2 protein of a strain of Chlamydia pneumoniae.  
CC Comparison of this sequence as to the recently published genome sequence  
CC of C. pneumoniae reveals that the sequence actually contains at least two  
CC open reading frames, a first one in the 5' portion and a second one in  
CC the 3' portion of the sequence. Despite the presence of the stop codon at  
CC the end of this sequence, C. pneumoniae does make a 76 kDa product. It  
CC appears possible that C. pneumoniae is able to read through this stop  
CC codon and produce a full-length product terminated by the stop codon at  
CC the end of the second open reading frame. There is at least one in-frame  
CC ATG upstream of the start codon. This suggests that the first open  
CC reading frame may form part of one or more larger open reading frames.  
CC The lorf2 protein or DNA can be used as a vaccine for humans to treat or  
CC prevent disease caused by Chlamydia infection. The sequences or an  
CC antibody to lorf2 can be used to diagnose a Chlamydia infection.  
XX  
SQ Sequence 1550 BP; 472 A; 268 C; 317 G; 493 T; 0 other;

Query Match 99.5%; Score 647.8; DB 21; Length 1550;  
Best Local Similarity 99.7%; Pred No. 3.2e-183;  
Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 atgacaaaaaacattatctgttggttgtagaaggattctcaatcgctttgcttaaacag 60  
Db 716 atgacaaaaaacattatctgttggttgtagaaggattctcaatcgctttgcttaaacag 775  
QY 61 tttttgtgaatgtagtgtgttcgactggaacacattcgttcctcagaaacccctccact 120  
Db 776 tttttgtgaatgtagtgtgttcgactggaacacattcgttcctcagaaacccctccact 835  
QY 121 acagaaaagctctacaacgctatgaaatacaatactgtttggcagtggtcgtc 180  
Db 836 acagaaaagctctacaacgctatgaaatacaatactgtttggcagtggtcgtc 895  
QY 181 ggaagcattagtcaggttctctggatcgaatggacagaaaagcctctatatctttatgga 240  
Db 896 ggaagcattagtcaggttctctggatcgaatggacagaaaagcctctatatctttatgga 955  
QY 241 gctttcttaataagacctttagcaagcctgttcagttccagaaatagatgtttcaggatt 300  
Db 956 gctttcttaataagacctttagcaagcctgttcagttccagaaatagatgtttcaggatt 1015  
QY 301 ctgacttggtttattgaggaacttttaggggactcagaaaagctggagactggctgccc 360  
Db 1016 ctgacttggtttattgaggaacttttaggggactcagaaaagctggagactggctgccc 1075  
QY 361 acagtcagttatgagtagtcgaagcctgttcagttccagaaatagatgtttcaggatt 420  
Db 1076 acagtcagttatgagtagtcgaagcctgttcagttccagaaatagatgtttcaggatt 1135  
QY 421 ggcctgtgtaattatttaaaagtgttgcgcacaaagcaattgctgctaactatgatcct 480  
Db 1136 ggcctgtgtaattatttaaaagtgttgcgcacaaagcaattgctgctaactatgatcct 1195  
QY 481 aaagagcctaagtagttttacaaaattataaaggattttccgctctatatgtatgtagc 540  
Db 1196 aaagagcctaagtagttttacaaaattataaaggattttccgctctatatgtatgtagc 1255  
QY 541 acagattctctatcattcagagcttatgggcttactccaaacagcaaacgataaactc 600  
Db 1256 acagattctctatcattcagagcttatgggcttactccaaacagcaaacgataaactc 1315  
QY 601 ggcagtgatttttactttccgaaaagtgttgatctaggtataatttcagcggttt 651  
Db 1316 ggcagtgatttttactttccgaaaagtgttgatctaggtataatttcagcggttt 1366

RESULT 2  
AAD02066  
ID AAD02066 standard; DNA; 2238 BP.  
XX  
AC AAD02066;  
XX  
DT 26-MAR-2001 (first entry)  
XX  
DE C. pneumoniae 76 kDa protein truncation mutant fusion gene.  
XX  
KW 76 kDa protein; bactericidal; diagnosis; prevention;  
KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;  
KW acute respiratory disease; cough; sore throat; hoarseness; fever;  
KW vaccine; immunisation; treatment; truncation mutant; fusion gene;  
KW ds.  
XX  
XX Chlamydia pneumoniae.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FT misc\_feature 1..665 /tag= a  
FT /note= "This part of the sequence is unrelated to  
FT C. pneumoniae 76 kDa gene"  
FT CDS 766..2238  
FT /tag= b  
FT /product= "Truncated Chlamydia pneumoniae  
FT 76kDa protein"  
FT misc\_feature 2122..2238  
FT /tag= c  
FT /note= "This part of the sequence is unrelated to  
FT C. pneumoniae 76 kDa gene"  
XX  
XX WO2000066739-A2.  
XX  
XX 09-NOV-2000.  
XX  
XX 03-MAY-2000; 2000WO-CA00511.  
XX  
XX 03-MAY-1999; 99US-0132270.  
XX  
XX 30-JUN-1999; 99US-0141276.  
XX  
XX (AVET ) AVENTIS PASTEUR LTD.  
XX  
XX Murdin AD, Oomen RP, Wang J, Dunn P;  
XX  
XX WPI: 2000-687542/67.  
XX  
XX P-PSDB; AAY71957.  
XX  
XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,  
XX useful for vaccinating against Chlamydia infections -  
XX  
XX Claim 32; Fig 3; 112pp; English.  
XX  
XX The present sequence is a DNA coding for a fusion protein comprising a  
XX truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded  
XX residues. C. pneumoniae 76 kDa protein is used in the  
XX diagnosis, prevention and treatment of C. pneumoniae  
XX infections (e.g. pneumonia, upper respiratory  
XX tract disease, bronchitis, sinusitis and acute respiratory  
XX disease such as cough, sore throat, hoarseness, fever; and  
XX abnormal chest sounds on auscultation). C. pneumoniae sequence  
XX is also used as vaccines for immunising humans against diseases  
XX caused by C. pneumoniae.  
XX  
XX Sequence 2238 BP; 689 A; 459 C; 497 G; 593 T; 0 other;  
SQ

Query Match 99.5%; Score 647.8; DB 21; Length 2238;  
Best Local Similarity 99.7%; Pred No. 3.7e-183;  
Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 atgacaaaaaacattatctgttggttgtagaaggattctcaatcgctttgcttaaacag 60



Db 1 atgacaaaaaacattatgcttgggtgtagaaggattctcaatcgtttgcttaaacag 60  
 Qy 61 ttttttggaaatagttgctgactggaacacattcgttctcagaaacccact 120  
 Db 61 ttttttggaaatagttgctgactggaacacattcgttctcagaaacccact 120  
 Qy 121 acagaaaaagctgctcaaacgctatgaatacaaaaactagtgtttggcagtgctgc 180  
 Db 121 acagaaaaagctgctcaaacgctatgaatacaaaaactagtgtttggcagtgctgc 180  
 Qy 181 ggaagcatagtcagggttccttgatcaatggacagaaaaagcctctatatctttatgga 240  
 Db 181 ggaagcatagtcagggttccttgatcaatggacagaaaaagcctctatatctttatgga 240  
 Qy 241 gctttcttaataaacctcttagcaaaagctacgaagactacgttaaatggaaaaaagaaac 300  
 Db 241 gctttcttaataaacctcttagcaaaagctacgaagactacgttaaatggaaaaaagaaac 300  
 Qy 301 ctactgtggtttattgagggaacttttaggggactcagaaaagctggagactggtctgcc 360  
 Db 301 ctactgtggtttattgagggaacttttaggggactcagaaaagctggagactggtctgcc 360  
 Qy 361 acagtacgtttatgagtagtgcgaagccttgctcagttccagaaatagatgtttcaggatt 420  
 Db 361 acagtacgtttatgagtagtgcgaagccttgctcagttccagaaatagatgtttcaggatt 420  
 Qy 421 ggcctgtgtaattattaaagtttgcgtcccaagcaattgctcgtctctatatgtatgacct 480  
 Db 421 ggcctgtgtaattattaaagtttgcgtcccaagcaattgctcgtctctatatgtatgacct 480  
 Qy 481 aaagagcctaattgctttacaaaataaaagattttccgcctctatatgtatgacct 540  
 Db 481 aaagagcctaattgctttacaaaataaaagattttccgcctctatatgtatgacct 540  
 Qy 541 acagattctctatcattcagagcttatgggcttactccaaaccagcaaaacgataaacctc 600  
 Db 541 acagattctctatcattcagagcttatgggcttactccaaaccagcaaaacgataaacctc 600  
 Qy 601 ggcagtgattttacttccgaaagttagatctaggtatataatttcagcgttt 651  
 Db 601 ggcagtgattttacttccgaaagttagatctaggtatataatttcagcgttt 651

RESULT 3  
 AAX91990/c  
 ID AAX91990 standard; DNA: 1230025 BP.  
 XX  
 AC AAX91990;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; ss.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffiths R;  
 XX

DR WPI; 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 XX  
 PS Claim 1; Page 291-611; 1912pp; English.  
 XX  
 CC The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.  
 XX  
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 99.5%; Score 647.8; DB 20; Length 1230025;  
 Best Local Similarity 99.7%; Pred. No. 4.5e-182;  
 Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 atgacaaaaaacattatgcttgggtgtagaaggattctcaatcgtttgcttaaacag 60  
 Db 829363 ATGACAAAAAACATTATGCTTGGTGTGACAGGGATTCTCAATCGTTTGCCTAAACAG 829304  
 Qy 61 ttttttggaaatagttgctgactggaacacattcgttctcagaaacccact 120  
 Db 829303 TTTTGTGTGAATAGTGTGCTGACTGGAACACATTCTGTTCTCAGAAACCTCCACT 829244  
 Qy 121 acagaaaaagctgctcaaacgctatgaatacaaaaactagtgtttggcagtgctgc 180  
 Db 829243 ACAGAAAAAGCTGCTACAAACCTATGAATACAAACTGTTGTTGCGCTGCTC 829184  
 Qy 181 ggaagcatagtcagggttccttgatcaatggacagaaaaagcctctatatgtatgga 240  
 Db 829183 GGAAAGCATAGTACAGTTCCTTGGATCAATGCAGAAAAAGCCTCTATATCTTTATGGA 829124  
 Qy 241 gctttcttaataaacctcttagcaaaagctacgaagactacgttaaatggaaaaaagaaac 300  
 Db 829123 GCTTTCCTTAATGAACCCCTTTAGCAAGCTACGTTAAATGGAAGAAAAA 829064  
 Qy 301 ctactgtggtttattgagggaacttttaggggactcagaaaagctggagactggtctgcc 360  
 Db 829063 CTAGCTTGGTTTATTGGAGGAACTTTAGGGGGACTCAGAAAAAGCTGGAGACTGCTCTGCC 829004  
 Qy 361 acagtacgtttatgagtagtgcgaagccttgctcagttccagaaatagatgtttcaggatt 420  
 Db 829003 ACAGTACGTTATGATGATGTCGAAAGCCTTGTGCTGTTCCAGAAATAGATGTTTCAGGGATT 828944  
 Qy 421 ggcctgtgtaattattaaagtttggctcccaagcaattcgtcgaactatgacct 480  
 Db 828943 GGCCGTGTAAATTATTAAAGTTTGTGTTGCCCAAGCAATGTCTGCTAACTATGATCCT 828884  
 Qy 481 aaagagcctaattgctttacaaaataaaagattttccgcctctatatgtatggaatc 540  
 Db 828883 AAAGAGGCTAATGCTTTTACAAATATATAAGGATTTTCCGCTCTATATATGATGCGATC 828824  
 Qy 541 acagattctctatcattcagagcttatgggcttactccaaaccagcaaaacgataaacctc 600  
 Db 828823 ACAGATTCTCTATCATTACAGAGCTTATGGGCTTACTCCAAACAGCAAGATAAATC 828764  
 Qy 601 ggcagtgattttacttccgaaagttagatctaggtatataatttcagcgttt 651  
 Db 828763 GGCAGTGATTTTACTTCCGAAAGTTTGTATCTAGGTATATAATTTCAGCGTTT 828713

RESULT 4  
 AAF58252/c  
 ID AAF58252 standard; DNA: 936 BP.  
 XX



















XX	26-JUL-2000; 2000WO-US20476.	
PF		
XX	PR 9905-0145695.	
XX	PR 2000US-0190259.	
XX	(CLIN-) CLINICAL MICRO SENSORS INC.	
XX		
XX	Umek RM;	
PI		
XX	WPI; 2001-159728/16.	
XX		
DR	Nucleic acids containing electron-transfer group, useful as labels in	
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on	
PT	a single surface	
XX		
XX	Example 6; Page 127; 159pp; English.	
XX		
CC	The present invention relates to a composition comprising two nucleic	
CC	acids each containing an electron-transfer group (ETM) having	
CC	different redox potentials. The invention is used for electronic	
CC	detection of nucleic acids, especially of substitutions (mismatches)	
CC	and single-nucleotide polymorphisms, e.g. for genotyping,	
CC	and monitoring gene expression.	
XX		
XX	Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;	
QS		
Query Match 13.0%; Score 84.6; DB 22; Length 936;		
Best Local Similarity 0.94; Pred. No. 2e-15;		
Matches 6; Conservative 384; Mismatches 253; Indels 0; Gaps		
Qy	1 atgcacacacacattatgcttgggttagaaggagattccaatcgtttgctctaaacag 60	
Db	136 www	195
Qy	61 ttttttgaaatgtagtgggtgcgactggagacacattgcttccttcgaaacctccact 120	
Db	196 gww	255
Qy	121 acagacaaaagctgctacaaacgctatgaaatacaatactgtgtggcagtggtcgc 180	
Db	256 www	315
Qy	181 ggaagcatagtcaggctccttgatcatgacagacaaagcctctatatctttatgga 240	
Db	316 www	375
Qy	241 gcttcttaatgaacccttttagcaaggctacgaagactacgtttaatggaaaagaaac 300	
Db	376 www	435
Qy	301 ctagcttggtttattgaggaaacttttagggggactcagaaaagctggagactggctgc 360	
Db	436 www	495
Qy	361 acagtaactatgagtatgtagaacctctgcagttccagaaatacatgtttcaggatt 420	
Db	496 www	555
Qy	421 ggcgtggtatattataaagtttggctgcgcacaaatgtgctgaactatgatct 480	
Db	556 www	615
Qy	481 aaagaggctaagtgttacaaattataaaggattttccgctctatatatgtatggcatc 540	
Db	616 www	675
Qy	541 acagattctctatcatcagacttatggggttactccaaacagacgaacgaacta 600	
Db	676 www	735
Qy	601 ggcagtgattttactttccgaaagtttgatcaggtataatt 643	

[illegible]



[illegible]

RESULT 12

RESOL 12  
AAAF58257  
ID AAAF58257 standard; DNA: 936 BP.

AA  
AC AAF58257:

DT 24-APR-2001 (first entry)

AA  
DE  
Oligonucleotide D1954.

Electron-transfer group; ETM; mismatch; genotyping;  
KW  
gene expression; ss.  
KW

XX  
OS  
Synthetic.

XX  
PN WO200107665-A2.

XX  
PD 01-FEB-2001.

26-JUL-2000: 2000WO-US20476.

XX 26-JUL-1999: 99US-0145695.

PK 20-JUL-1999; 9903-0143093;  
PR 17-MAR-2000; 2000US-0190259;

PA (CLIN-) CLINICAL MICRO SENSORS INC.

XX  
PI  
Umek RM:

AA  
DR  
WPI; 2001-159728/16.

AA Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
PT

XX  
PS  
Example 6: Page 127: 159pp: English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match  
13.0%; Score 84.6; DB 22; Length 936;

Best Local Similarity 0.9%; Pred. No. 2e-15;  
Matches 6; Conservative 384; Mismatches 253; Indels 0; Gaps 0;

[illegible]

RESULT 13

RESULT 13  
AAAF58259  
ID AAAF58259 standard: DNA: 936 BP.

XX AC "A" F58259:

XX  
DT 24-APR-2001 (first entry)

XX DE Oligonucleotide D2004.

XX Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression: ss.

XX OS Synthetic.

XX PN WO200107665-A2.

XX PD 01-FEB-2001.

XX  
PF 26-JUL-2000; 2000WO-US20476.

XX  
PR 26-JUL-1999; 99US-0145695.

PR 20-JUL-1999; 990US-0143093;  
PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.



XX	AAF58262;
AC	
XX	24- APR-2001 (first entry)
DT	
XX	Oligonucleotide D2007.
DE	
XX	
KW	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
XX	
XX	Synthetic.
OS	
XX	WC200107665-A2.
PN	
XX	
PD	01-FEB-2001.
XX	
XX	26-JUL-2000; 2000WO-US20476.
PF	
XX	26-JUL-1999; 99US-0145695.
PR	

PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	
PI	Umek RM;
XX	
DR	WPI; 2001-159728/16.
XX	
PT	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface -
XX	
PS	Example 6; Page 128; 159pp; English.
XX	
CC	The present invention relates to a composition comprising two nucleic
CC	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
SQ	Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;
Query Match	13.0%; Score 84.6; DB 22; Length 936;
Best Local Similarity	0.9%; Pred. No. 28-15;
Matches	6; Conservative 384; Mismatches 253; Indels 0; Gaps
Qy	1 atgcaaaaaaacattatgtcgggttgtagaaggattctcaatcgttgcctaacaag 60 : Db 136 www. ....
Qy	61 ttttltgaaatgtagtgttcgactggaacacattcgttccttcagaacctccact 120 : Db 196 gwww. ....
Qy	121 acagaaaaagctgctacaaacgctatgaaatacacaaactcgttgttggaactgctc 180 : Db 256 www. ....
Qy	181 ggaagcatagtcaggttccttggatccaatggacagaaaaagcctctattcttggga 240 : Db 316 www. ....
Qy	241 gctttttaatgaacccttagcaaaaggctacgaagactacgtttaaatggaaaagaac 300 : : : : : : :     : Db 376 www. ....
Qy	301 ctagcttggtttattggaagaaacttaggggactcagaaaaagctggagactggctgcc 360

Qy 361 acagtacgttatgagtagtgcgaagccttgcagttccagaatatgattgttccaggatt 420  
430



Db 496 ..... 555  
 QY 421 ggcctgtaatttattaaagtttgggtgcgccaagcaattgctgtaactatgatcct 480  
 Db 556 ..... 615  
 QY 481 aaagaggctaatagtttacaaattataaaagattttccgctctataatgatgcatc 540  
 Db 616 ..... 675  
 QY 541 acagattctctatcattcagagcttatgggcttactccaaaccgcaaacgataaac 600  
 Db 676 ..... 735  
 QY 601 ggcagtgatttacttccgaaagtgtgatctaggtataatt 643  
 Db 736 ..... 778

RESULT 15  
 AAF58255  
 ID AAF58255 standard; DNA; 938 BP.  
 AC AAF58255;  
 XX 24-APR-2001 (first entry)  
 XX Oligonucleotide D1876.  
 XX  
 KW Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200107665-A2.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 26-JUL-2000; 200WO-US20476.  
 XX  
 PR 26-JUL-1999; 99US-0145695.  
 PR 17-MAR-2000; 2000US-0190259.  
 XX  
 PA (CLIN-) CLINICAL MICRO SENSORS INC.  
 XX  
 PI Umek RM;  
 XX  
 DR WPI; 2001-159728/16.  
 XX  
 PT Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX  
 PS Example 6; Page 127; 159pp; English.  
 XX  
 CC The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX  
 SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

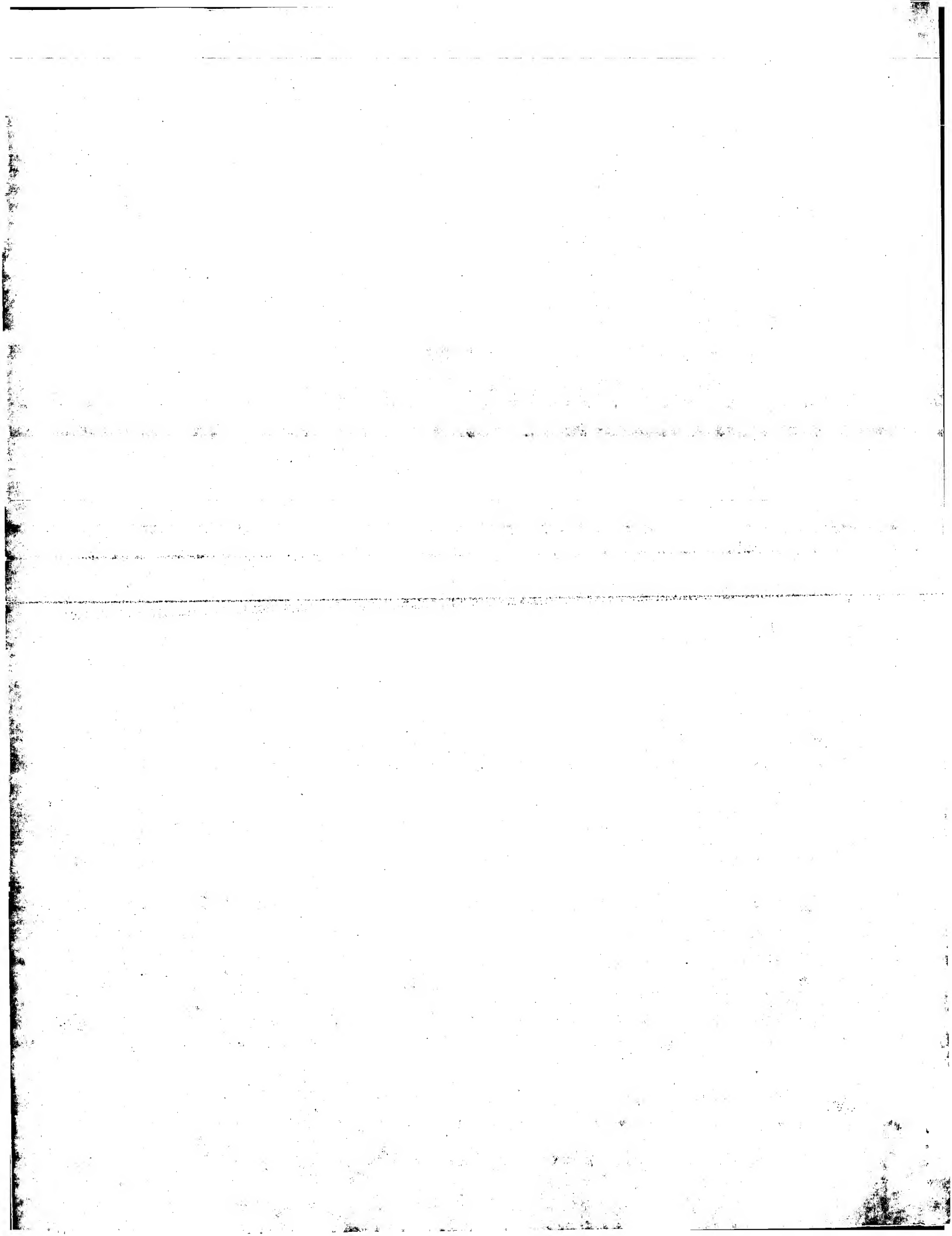
Query Match 13.0%; Score 84.6; DB 22; Length 938;  
 Best Local Similarity 0.9%; Pred No. 2e-15;  
 Matches 6; Conservative 384; Mismatches 253; Indels 0; Gaps 0;

QY 1 atgacaaaaaacattgcttggttagaaggattctcaatcggttgctaaacag 60  
 Db 136 ..... 195

Search completed: February 7, 2002, 19:50:20  
 Job time: 14534 sec

QY 61 ttttttgtaagttagttgctgactggaaacacattcgttcttcagaaaacctccact 120  
 Db 196 g..... 255  
 QY 121 acagaaaagctgctacaaacgctatgaaatcacaaatactgtgttgagcagtggtcgtc 180  
 Db 256 ..... 315  
 QY 181 gaaagcagtagtcaggttccttgatcaatggacagaaagcctctatatcttatgga 240  
 Db 316 ..... 375  
 QY 241 gctttcttaatgaacccotttagcaagctacagagctaaagttaattggaagaaac 300  
 Db 376 ..... 435  
 QY 301 ctgcttggtttattgaggaacttttaggggactcagaaaagctggagactggtctgcc 360  
 Db 436 ..... 495  
 QY 361 acagtcatttatgagtatgtcgaagccttgctcagttccagaaatagatgtttcaggatt 420  
 Db 496 ..... 555  
 QY 421 ggcctggttaattattaaagtttgggtgcgccaagcaattgctgtaactatgatcct 480  
 Db 556 ..... 615  
 QY 481 aaagaggctaatagtttacaaattataaaagattttccgctctataatgatgcatc 540  
 Db 616 ..... 675  
 QY 541 acagattctctatcattcagagcttatgggcttactccaaaccgcaaacgataaac 600  
 Db 676 ..... 735  
 QY 601 ggcagtgatttacttccgaaagtgtgatctaggtataatt 643  
 Db 736 ..... 778







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:53:24 ; Search time 243.49 Seconds  
(without alignments)  
605.516 Million cell updates/sec

Title: US-09-391-606-2

Perfect score: 651

Sequence: 1 atgacaaaaaacattatgc.....taggtataatttcagcggttt 651

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	36.6	5.6	15225	2	US-08-892-403A-2
2	35.6	5.5	3208	2	US-08-566-398-51
3	35.6	5.5	3209	1	US-08-105-483-220
4	35.6	5.5	3209	1	US-08-220-151-62
5	35.6	5.5	3209	1	US-08-413-118-62
6	35.6	5.5	3209	1	US-08-224-391-90
7	35.6	5.5	3209	1	US-08-484-304-90
8	35.6	5.5	3209	1	US-08-224-657-39
9	35.6	5.5	3209	1	US-08-709-209-220
10	35.6	5.5	3209	1	US-08-458-101-220
11	35.6	5.5	3209	2	US-08-184-009-68
12	35.6	5.5	3209	2	US-08-486-969-27
13	35.6	5.5	3209	2	US-08-417-210A-66
14	35.6	5.5	3209	2	US-08-458-356-68
15	35.6	5.5	3209	2	US-08-471-025-27
16	35.6	5.5	3209	2	US-08-658-665-27
17	35.6	5.5	3209	3	US-08-473-446-62
18	35.6	5.5	3209	4	US-08-796-101-43
19	35.6	5.5	3209	4	US-08-460-736-68
20	35.6	5.5	3209	4	US-09-085-273-27
21	35.6	5.5	3209	5	PCT-US96-00547-27
22	33.6	5.2	1894	2	US-08-935-450-7
c 23	32.6	5.0	19011	1	US-08-310-356-36
c 24	32.6	5.0	19557	5	PCT-US92-06300-1
25	31.6	4.9	3077	1	US-08-703-809-1
26	31.6	4.9	3077	1	US-08-703-808-1
27	31.6	4.9	3077	2	US-08-914-066-1

28	31.6	4.9	3077	2	US-08-703-807-1	Sequence 1, Appli
29	31.6	4.9	3077	2	US-08-747-108A-1	Sequence 1, Appli
30	31.6	4.9	3077	3	US-09-211-631-1	Sequence 1, Appli
31	31.6	4.9	3077	3	US-09-265-628-1	Sequence 1, Appli
32	31.6	4.9	3077	4	US-09-001-141-1	Sequence 1, Appli
33	31.6	4.9	3077	4	US-09-532-803-1	Sequence 1, Appli
c 34	31	4.8	2433	1	US-08-136-743B-40	Sequence 40, Appli
c 35	31	4.8	2663	1	US-08-136-743B-3	Sequence 3, Appli
c 36	31	4.8	19124	2	US-08-487-826B-13	Sequence 13, Appli
37	30.8	4.7	1749	1	US-08-116-098-1	Sequence 1, Appli
38	30.6	4.7	5319	1	US-08-169-927-1	Sequence 1, Appli
c 39	30.6	4.7	7218	1	US-08-232-463-14	Sequence 14, Appli
40	30.4	4.7	2791	2	US-09-016-000-8	Sequence 8, Appli
41	30.2	4.6	1770	3	US-09-006-636-1	Sequence 1, Appli
42	30.2	4.6	1770	4	US-09-006-632-1	Sequence 1, Appli
43	30.2	4.6	2687	3	US-08-714-918-34	Sequence 34, Appli
44	30.2	4.6	2687	4	US-09-265-315-34	Sequence 34, Appli
45	30.2	4.6	2687	4	US-09-265-315-34	Sequence 34, Appli

## ALIGNMENTS

### RESULT 1

US-08-892-403A-2/c  
Sequence 2, Application US/08892403A

Patent No. 5993824

GENERAL INFORMATION:

APPLICANT: Murphy, Brian R.

APPLICANT: Collins, Peter L.

APPLICANT: Whitehead, Stephen S.

APPLICANT: Bukreyev, Alexander A.

APPLICANT: Juhasz, Katalin

TITLE OF INVENTION: PRODUCTION OF ATTENUATED RESPIRATORY

TITLE OF INVENTION: SYNCTIAL VIRUS VACCINES FROM CLONED NUCLEOTIDE SEQUENCES

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/892.403A

FILING DATE: 15-JUL-1997

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/047.634

FILING DATE: 23-MAY-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/046.141

FILING DATE: 09-MAY-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/021.773

FILING DATE: 15-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: Parmelee, Steven W.

REGISTRATION NUMBER: 31,990

REFERENCE/DOCKET NUMBER: 17634-000510

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-467-9600

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 15225 base pairs

TYPE: nucleic acid

STRANDEDNESS: single



TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-892-403A-2

Query Match 5.6%; Score 36.6; DB 2; Length 15225;  
Best Local Similarity 55.0%; Pred. No. 0.13; 59; Indels 0; Gaps 0;  
Matches 72; Conservative 0; Mismatches 0;

QY 405 agatgtttcaggagtgccggtggaatttattataaagtttgggttcgcccgaagcaattgc 464  
DB 8397 AGATGTTGGAGAAATGTTGAGTGCATCTAATAAGTTTGTAGGTGCCAATGAATTCA 8338  
QY 465 tgcatactatgatctctaagaggctaatagttttacaaaattataaaggattttccgctct 524  
DB 8337 TCAAAATATATATTTTAAATAAAGCGTGGTGTATCCAAATGTTATGCTTAAATGTAAT 8278

QY 525 atatgtatg 535  
DB 8277 ATGTTTTTAIG 8267

RESULT 2

US-08-566-398-51  
Sequence 51, Application US/08566398  
Patent No. 5658373

GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
TITLE OF INVENTION: RECOMBINANT POXVIRUS - FELINE INFECTIONS  
TITLE OF INVENTION: PERITONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR  
TITLE OF INVENTION: MAKING AND USING THEM  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/566,398  
FILING DATE: 01-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2880  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3208 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
US-08-566-398-51  
Query Match 5.5%; Score 35.6; DB 2; Length 3208;  
Best Local Similarity 49.6%; Pred. No. 0.13;  
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

QY 348 agactggtctgcacagtcagtcagtgtatgtcgaagccttgcagttccagaataaga 407  
DB 2446 AGATAGAGGGCCAAAGTAATTAACAGGATGTTACGGTCTTCAGCTCTCCATAGAGC 2505

QY 408 tgtttcaggagtgccggtggaatttattataaagtttgggttcgcccgaagcaattgtcgc 467  
DB 2506 TGCTATTGTAGGAACAGG--ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2562  
QY 468 taactatgatcctaaagagcgttaagttttacaaaattataaaggattttccgctctata 527  
DB 2563 AAACCTTTAACTATTGCTTAAGATAATCTTATTAATAAATAATAATATCATCAGTTTAGTA 2622  
QY 528 tatgtatgcatacacagattctctatcattcacagcttatgggcttactccaaccagc 587  
DB 2623 ATATTAATAATATATTAATAACTCTTATTACTTAATAACTCCAGTGGATATGAACATAATACG 2682  
QY 588 aa 589  
DB 2683 AA 2684

RESULT 3

US-08-105-483-220  
Sequence 220, Application US/08105483  
Patent No. 5494807

GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE  
TITLE OF INVENTION: STRAIN  
NUMBER OF SEQUENCES: 462  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
ADDRESSEE: c/o William S. Frommer  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/105,483  
FILING DATE: 12-AUG-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/847,951  
FILING DATE: 06-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2400  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 220:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3209 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-105-483-220  
Query Match 5.5%; Score 35.6; DB 1; Length 3209;  
Best Local Similarity 49.6%; Pred. No. 0.13; 119; Indels 3; Gaps 1;  
Matches 120; Conservative 0; Mismatches 119;

QY 348 agactggtctgcacagtcagtcagtgtatgtcgaagccttgcagttccagaataaga 407  
DB 2447 AGATAGAGGGCCAAAGTAATTAACAGGATGTTTACGCTTCTTCAGCTCTCCATAGAGC 2506

QY 408 tgtttcaggagtgccggtggaatttattataaagtttgggttcgcccgaagcaattgtcgc 467



Db 2507 TCGTATTGGTAGAAACAGG---ATATGATAAAGCTGTTAATCGCATCATGGAGCTGTGT 2563  
Qy 468 taactatgatctctaagagcctaagtagtttacaaattataaagattttccgctctata 527  
Db 2564 AACTCTTTAACTATTGCTAAAGATAAATCTATTAAAAAATAATATACGTTTAGTA 2623  
Qy 528 tatgtatggatcacagattctctatctatcattcagagcttatgggcttactccaaccagc 587  
Db 2624 ATATTAAAAATATTAATAAAGCTATTACTAATAACTCCAGTGGATGATGAACATAATACG 2683  
Qy 588 aa 589  
Db 2684 AA 2685

## RESULT 4

US-08-220-151-62  
; Sequence 62, Application US/08220151  
; Patent No. 5529780  
; GENERAL INFORMATION:  
; APPLICANT: Paolletti, Enzo  
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
; TITLE OF INVENTION: CANINE HERPESVIRUS 9B, 9C AND 9D AND USES THEREFOR  
; NUMBER OF SEQUENCES: 91  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/220,151  
; FILING DATE: 30-MAR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer, William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454310-2540  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; TELEX: 425066 CURTMS  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3209 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-220-151-62

Query Match 5.5%; Score 35.6; DB 1; Length 3209;  
Best Local Similarity 49.6%; Pred. No. 0.13;  
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;  
Qy 348 agactggtctgccacagtcacgttatgatgtcgaagccttgcagttccagaaataga 407  
Db 2447 AGATAGAGGGGCCAAAGTAATAATACAGGATGTTTACGGTCTTTCAGCTCTCCATAGAGC 2506  
Qy 408 ttttccaggattggccgtggttaatttataaagtttgggttcgcccagaagaattgcgc 467  
Db 2507 TCGTATTGGTAGAAACAGG---ATATGATAAAGCTGTTAATCGATCATGAGCTGTGT 2563  
Qy 468 taactatgatctctaagagcctaagtagtttacaaattataaagattttccgctctata 527

Db 2564 AACTCTTTAACTATTGCTAAAGATAAATCTTTAAAAAATAATATACGTTTAGTA 2623  
Qy 528 tatgtatggatcacagattctctatcattcagagcttatgggcttactccaaccagc 587  
Db 2624 ATATTAAAAATATTAATAAAGCTATTACTAATAACTCCAGTGGATGATGAACATAATACG 2683  
Qy 588 aa 589  
Db 2684 AA 2685

## RESULT 5

US-08-413-118-62  
; Sequence 62, Application US/08413118  
; Patent No. 5688920  
; GENERAL INFORMATION:  
; APPLICANT: PAOLETTI, ENZO  
; APPLICANT: LIMBACH, KEITH J.  
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
; TITLE OF INVENTION: CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR  
; NUMBER OF SEQUENCES: 128  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
; STREET: 530 FIFTH AVENUE, 25TH FLOOR  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/413,118  
; FILING DATE: 29-MAR-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/220,151  
; FILING DATE: 30-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FROMMER, WILLIAM S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454310-2670  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3209 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-413-118-62

Query Match 5.5%; Score 35.6; DB 1; Length 3209;  
Best Local Similarity 49.6%; Pred. No. 0.13;  
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;  
Qy 348 agactggtctgccacagtcacgttatgatgtcgaagccttgcagttccagaaataga 407  
Db 2447 AGATAGAGGGGCCAAAGTAATAATACAGGATGTTTACGGTCTTTCAGCTCTCCATAGAGC 2506  
Qy 408 ttttccaggattggccgtggttaatttataaagtttgggttcgcccagaagaattgcgc 467  
Db 2507 TCGTATTGGTAGAAACAGG---ATATGATAAAGCTGTTAATCGATCATGAGCTGTGT 2563  
Qy 468 taactatgatctctaagagcctaagtagtttacaaattataaagattttccgctctata 527  
Db 2564 AACTCTTTAACTATTGCTAAAGATAAATCTTTAAAAAATAATATACGTTTAGTA 2623











```

RESULT 10
US-08-458-101-220
: Sequence 220, Application US/08458101
: Patent No. 5766599
: GENERAL INFORMATION:
: APPLICANT: Paoletti, Enzo
: APPLICANT: Perkus, Marion E.
: APPLICANT: Taylor, Jill
: APPLICANT: Tartaglia, James
: APPLICANT: No. 5766599ton, Elizabeth K.
: APPLICANT: Riviere, Michel
: APPLICANT: de Taisne, Charles
: APPLICANT: Limbach, Keith J.
: APPLICANT: Johnson, Gerard P.
: APPLICANT: Pincus, Steven E.
: APPLICANT: Cox, William I.
: APPLICANT: Audonnet, Jean-Christophe Francis
: APPLICANT: Gettig, Russell Robert
: TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE
: TITLE OF INVENTION: STRAIN
: NUMBER OF SEQUENCES: 467
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Curtis, Morris & Safford
: ADDRESSEE: c/o William S. Frommer
: STREET: 530 Fifth Avenue
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,101
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Frommer, William S.
: REGISTRATION NUMBER: 25,506
: REFERENCE/DOCKET NUMBER: 454310-2740
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 840-3333
: TELEFAX: (212) 840-0712
: INFORMATION FOR SEQ ID NO: 220:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3209 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-458-101-220

Query Match 5.5%; Score 35.6; DB 1; Length 3209;
Best Local Similarity 49.6%; Pred. No. 0.13;
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

QY 348 agactggtctgccacagtcagttatgagtcgtcgaagccttgctcagttccagaaataga 407
Db 2447 AGATAGAGGGGCCAAAGTAATACAGAGGATGTTTACGGTTCTTCAGCTCTCCATAGAGC 2506
QY 408 tgtttcaggagattggccgtggtaatttataaagtttgggttcgcccagaacatttctgc 467
Db 2507 TGCTATTGGTAGAAGACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563
QY 468 taactatgactcacaagagcctaattttacaaataataaagattttccgcctctata 527
Db 2564 AACTCTTTAACTATTGCTAAAGATAATCTTATTAATAAATAATAATATACGTTTAGTA 2623
QY 528 tatgtatggcatcacagattctctatcattcagagcttatgggcttactccaaccagc 587
Db 2624 ATATTAATAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2683
QY 588 aa 589

US-08-184-009-68
: Sequence 68, Application US/08184009
: Patent No. 5833975
: GENERAL INFORMATION:
: APPLICANT: Paoletti, Enzo
: APPLICANT: Tartaglia, James
: APPLICANT: Cox, William I.
: TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
: NUMBER OF SEQUENCES: 217
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Curtis, Morris & Safford
: STREET: 530 Fifth Avenue
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/184,009
: FILING DATE: 19-JAN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Frommer, William S.
: REGISTRATION NUMBER: 25,506
: REFERENCE/DOCKET NUMBER: 454310-2530
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 840-3333
: TELEFAX: (212) 840-0712
: TELEX: 425066CURTMS
: INFORMATION FOR SEQ ID NO: 68:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3209 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-184-009-68

Query Match 5.5%; Score 35.6; DB 2; Length 3209;
Best Local Similarity 49.6%; Pred. No. 0.13;
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

QY 348 agactggtctgccacagtcagttatgagtcgtcgaagccttgctcagttccagaaataga 407
Db 2447 AGATAGAGGGGCCAAAGTAATACAGAGGATGTTTACGGTTCTTCAGCTCTCCATAGAGC 2506
QY 408 tgtttcaggagattggccgtggtaatttataaagtttgggttcgcccagaacatttctgc 467
Db 2507 TGCTATTGGTAGAAGACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563
QY 468 taactatgactcacaagagcctaattttacaaataataaagattttccgcctctata 527
Db 2564 AACTCTTTAACTATTGCTAAAGATAATCTTATTAATAAATAATAATATACGTTTAGTA 2623
QY 528 tatgtatggcatcacagattctctatcattcagagcttatgggcttactccaaccagc 587
Db 2624 ATATTAATAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2683
QY 588 aa 589
```



Db 2684 AA 2685

## RESULT 12

US-08-486-969-27

: Sequence 27, Application US/08486969

: Patent No. 5843456

: GENERAL INFORMATION:

: APPLICANT: Paolletti, Enzo

: TITLE OF INVENTION: RECOMBINANT POXVIRUS - RABIES

: TITLE OF INVENTION: COMPOSITIONS AND COMBINATION COMPOSITIONS AND USES

: NUMBER OF SEQUENCES: 55

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Curtis, Morris &amp; Safford, P.C.

: STREET: 530 Fifth Avenue, 25th Floor

: CITY: New York

: STATE: New York

: COUNTRY: United States of America

: ZIP: 10036

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent in Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/486,969

: FILING DATE: 07-JUN-1995

: CLASSIFICATION: 424

: ATTORNEY/AGENT INFORMATION:

: NAME: Frommer, William S.

: REGISTRATION NUMBER: 25,506

: REFERENCE/DOCKET NUMBER: 454310-2600

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (212) 840-3333

: TELEFAX: (212) 840-0712

: INFORMATION FOR SEQ ID NO: 27:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 3209 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: cDNA

US-08-486-969-27

Query Match 5.5%; Score 35.6; DB 2; Length 3209;

Best Local Similarity 49.6%; Pred. No. 0.13;

Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

Qy 348 agactggtctgccacagtcacgttatgagtcagtcgaagcccttgctcagttccagaaataga 407

Db 2447 AGATAGAGGGGCCAAAGTAATTAACAAGGATGTTTACGGTCTTCAGCTCTCCATAGAGC 2506

Qy 408 tggttcaggagattggccgtggttaatttataaagtttgggttcgccccagcaattgctgc 467

Db 2507 TGCATTGTTAGGAAACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563

Qy 468 taactatgctcctaaagagcgaatagttttacaaattataaagattttccgctctata 527

Db 2564 AAACCTTTAACTATTGCTAAAGATAATCTTATTAAAAAATAATATCAGCTTAGTA 2623

Qy 528 tatgtatggcctacagattctctatcattcagagcttatgggcttactcctcaaccagc 587

Db 2624 ATATTAAATATATTATAAATCTTATTACTTAATAACTCCAGTGGATATGAACATAATAGC 2683

Qy 588 aa 589

Db 2684 AA 2685

## RESULT 13

US-08-417-210A-66

: Sequence 66, Application US/08417210A

: Patent No. 5863542

: GENERAL INFORMATION:

: APPLICANT: PAOLETTI, ENZO

: APPLICANT: TARTAGLIA, JAMES

: APPLICANT: COX, WILLIAM I.

: TITLE OF INVENTION: IMMUNODEFICIENCY RECOMBINANT POXVIRUS

: NUMBER OF SEQUENCES: 148

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: CURTIS, MORRIS &amp; SAFFORD, P.C.

: STREET: 530 FIFTH AVENUE

: CITY: NEW YORK

: STATE: NEW YORK

: COUNTRY: USA

: ZIP: 10036

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent in Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/417,210A

: FILING DATE: 05-APR-1995

: CLASSIFICATION: 435

: ATTORNEY/AGENT INFORMATION:

: NAME: KOWALSKI, THOMAS J.

: REGISTRATION NUMBER: 32,147

: REFERENCE/DOCKET NUMBER: 454310-2690

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 212-840-3333

: INFORMATION FOR SEQ ID NO: 66:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 3209 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

US-08-417-210A-66

Query Match 5.5%; Score 35.6; DB 2; Length 3209;

Best Local Similarity 49.6%; Pred. No. 0.13;

Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

Qy 348 agactggtctgccacagtcacgttatgagtcagtcgaagcccttgctcagttccagaaataga 407

Db 2447 AGATAGAGGGGCCAAAGTAATTAACAAGGATGTTTACGGTCTTCAGCTCTCCATAGAGC 2506

Qy 408 tggttcaggagattggccgtggttaatttataaagtttgggttcgccccagcaattgctgc 467

Db 2507 TGCATTGTTAGGAAACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563

Qy 468 taactatgctcctaaagagcgaatagttttacaaattataaagattttccgctctata 527

Db 2564 AAACCTTTAACTATTGCTAAAGATAATCTTATTAAAAAATAATATCAGCTTAGTA 2623

Qy 528 tatgtatggcctacagattctctatcattcagagcttatgggcttactcctcaaccagc 587

Db 2624 ATATTAAATATATTATAAATCTTATTACTTAATAACTCCAGTGGATATGAACATAATAGC 2683

Qy 588 aa 589

Db 2684 AA 2685

## RESULT 14

US-08-458-356-68

: Sequence 68, Application US/08458356

: Patent No. 5942235

: GENERAL INFORMATION:

: APPLICANT: Paolletti, Enzo

: APPLICANT: Tartaglia, James



APPLICANT: Cox, William I.  
 TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY  
 NUMBER OF SEQUENCES: 217  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Curtis, Morris & Safford  
 STREET: 530 Fifth Avenue  
 CITY: New York  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/458,356  
 FILING DATE: 02-JUN-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/184,009  
 FILING DATE: 19-JAN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Frommer, William S.  
 REGISTRATION NUMBER: 25,506  
 REFERENCE/DOCKET NUMBER: 454310-2530  
 TELEPHONE: (212) 840-3333  
 TELEFAX: (212) 840-0712  
 TELEX: 425066CURTMS  
 INFORMATION FOR SEQ ID NO: 68:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3209 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 US-08-458-356-68

Query Match 5.5%; Score 35.6; DB 2; Length 3209;  
 Best Local Similarity 49.6%; Pred. No. 0.13;  
 Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;  
 QY 348 agactggtctccacagtcacgtatgtatgagtcgagccttgcagttccagaaataga 407  
 DB 2447 AGATAGAGGGGCCAAAGTAATACAGGATGTTTACGGTCTTCAGCTCCATAGAGC 2506  
 QY 408 tgtttcagggttgccggtgtaattattaaagtgttgcgcccagcaattgctgc 467  
 DB 2507 TGCTATTGGTAGGAACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563  
 QY 468 taactatgatcctaaaggagcctaagttttacaaattataaaggattttccgctctata 527  
 DB 2564 AAACCTCTTAACCTATTGCTAAAGATAATCTTATTAAAAAATAATATCATCGTTTAGTA 2623  
 QY 528 tatgtatggcatcacagattcttcattcagagcttatgggcttactccaaaccagc 587  
 DB 2624 ATATTAAATATATTAACTATCTATTACTTAATCTCCAGTGGGATATGACATAATACG 2683  
 QY 588 aa 589  
 DB 2684 AA 2685

RESULT 15  
 US-08-471-025-27  
 ; Sequence 27; Application US/08471025  
 ; Patent No. 5989561  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Paolletti, Enzo  
 ; APPLICANT: Fischer, Laurent  
 ; APPLICANT: Legros, Francois-Xavier

TITLE OF INVENTION: RECOMBINANT POXVIRUS - CALICIVIRUS  
 TITLE OF INVENTION: [RABBIT HEMORRHAGIC DISEASE VIRUS (RDHV)] COMPOSITIONS AND  
 TITLE OF INVENTION: USES  
 NUMBER OF SEQUENCES: 47  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Curtis, Morris & Safford, P.C.  
 STREET: 530 Fifth Avenue, 25th Floor  
 CITY: New York  
 STATE: New York  
 COUNTRY: United States of America  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/471,025  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Frommer, William S.  
 REGISTRATION NUMBER: 25,506  
 REFERENCE/DOCKET NUMBER: 454310-2650  
 TELEPHONE: (212) 840-3333  
 TELEFAX: (212) 840-0712  
 INFORMATION FOR SEQ ID NO: 27:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3209 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 US-08-471-025-27

Query Match 5.5%; Score 35.6; DB 2; Length 3209;  
 Best Local Similarity 49.6%; Pred. No. 0.13;  
 Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;  
 QY 348 agactggtctccacagtcacgtatgtatgagtcgagccttgcagttccagaaataga 407  
 DB 2447 AGATAGAGGGGCCAAAGTAATACAGGATGTTTACGGTCTTCAGCTCCATAGAGC 2506  
 QY 408 tgtttcagggttgccggtgtaattattaaagtgttgcgcccagcaattgctgc 467  
 DB 2507 TGCTATTGGTAGGAACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563  
 QY 468 taactatgatcctaaaggagcctaagttttacaaattataaaggattttccgctctata 527  
 DB 2564 AAACCTCTTAACCTATTGCTAAAGATAATCTTATTAAAAAATAATATCATCGTTTAGTA 2623  
 QY 528 tatgtatggcatcacagattcttcattcagagcttatgggcttactccaaaccagc 587  
 DB 2624 ATATTAAATATATTAACTATCTATTACTTAATCTCCAGTGGGATATGACATAATACG 2683  
 QY 588 aa 589  
 DB 2684 AA 2685

Search completed: February 7, 2002, 15:53:37  
 Job time: 331 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 21:32:10 ; Search time 7419.31 Seconds  
(without alignments)  
942.878 Million cell updates/sec

Title: US-09-391-606-2  
Perfect score: 651  
Sequence: 1 atgacaaaaaacattatgc.....taggtataatttcagcggttt 651

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rpd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	6.0	474	10	AW569587
2	38.2	5.9	498	13	AQ769224
3	37.8	5.8	1894	12	AK013338
4	37.6	5.8	446	13	AZ933264
5	37.2	5.7	294	10	AB173567
6	37	5.7	303	10	AA184415
7	37	5.7	412	10	AA185888
8	37	5.7	531	13	AZ017765
9	36.8	5.7	780	10	AA563513
10	36.8	5.7	1101	13	CNS017V6
11	36.6	5.6	914	13	CNS0077X
12	36.6	5.6	1101	13	CNS0039G

13	35.4	5.6	1101	13	CNS0106X
14	35.2	5.6	555	13	AZ324640
15	35.2	5.6	846	13	AZ689918
16	36	5.5	694	11	BG440127
17	35.8	5.5	419	10	AU086798
18	35.8	5.5	660	10	AW040929
19	35.8	5.5	871	13	CNS04TMN
20	35.8	5.5	1101	13	CNS00JJR
21	35.6	5.5	375	10	BE185305
22	35.6	5.5	643	13	AZ977355
23	35.4	5.4	484	13	AZ117341
24	35.4	5.4	539	11	BF840086
25	35.2	5.4	333	10	AV206371
26	35.2	5.4	475	13	AQ086655
27	35.2	5.4	539	13	AQ751582
28	35.2	5.4	733	13	AZ945029
29	35	5.4	540	13	AQ514730
30	35	5.4	571	13	AZ066718
31	35	5.4	625	10	AV398733
32	35	5.4	1101	13	CNS00DQU
33	34.8	5.3	272	10	BE080531
34	34.8	5.3	565	13	AZ483477
35	34.8	5.3	713	10	BE661144
36	34.8	5.3	880	13	AZ683973
37	34.8	5.3	899	13	AZ538849
38	34.8	5.3	1101	13	CNS017KX
39	34.6	5.3	501	13	AQ777320
40	34.6	5.3	1086	13	CNS00YXK
41	34.4	5.3	341	10	AI587863
42	34.4	5.3	623	13	AZ872370
43	34.4	5.3	679	13	AZ638237
44	34.4	5.3	772	10	BE314620
45	34.4	5.3	834	13	CNS03EF2

ALIGNMENTS

RESULT 1  
AW569587  
LOCUS  
DEFINITION  
Gm-c1031-1820 5' similar to SW:TCQP\_HUMAN P50990 T-COMPLEX PROTEIN  
1, THETA SUBUNIT ;, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Glycine max  
soybean.  
REFERENCE  
AUTHORS  
Shoemaker,R., Kelm,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna  
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk  
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann  
,R., Waterston,R. and Willson,R.  
Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or



info@genomesystems.com web site: www.genomesystems.com  
Insert length: 700 Std Error: 0.00  
High quality sequence stop: 436.  
Location/Qualifiers

FEATURES

source

1. .474  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1031-1820"  
/clone\_lib="Gm-c1031"  
/tissue\_type="Williams seedlings, minus the cotyledons"  
/lab\_host="DH10B"  
/note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2: XhoI; This cDNA library was constructed from mRNA isolated from whole 'Williams' seedlings, minus the cotyledons, which were propagated on paper towels with distilled water for 5 days, incubated at 40 degrees C for 1 hour. The cotyledons were removed and the remaining tissue was flash-frozen in liquid nitrogen. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An 'anchor' nucleotide (V-A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGACTAGTCGAC(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second strand synthesis, the cDNA ends were filled in with cloned pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-, DNase-free water. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI from Promega (400/ul); all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500 bp cutoff, using GibcoBRL Life Technologies' cDNA Size fractionation column. The column eluent was then precipitated, redissolved, and ligated into Stratagene's pBluescript II XR predigested vector (pBluescript II SK(+)) vector that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). 100% of the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=18 and 5, respectively). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."  
BASE COUNT 136 a 127 g 133 t  
ORIGIN

Query Match 6.0%; Score 39; DB 10; Length 474;  
Best Local Similarity 50.8%; Pred. No. 2.7;  
Matches 93; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
Qy 228 atatcttttgagcgtttttaaataacccttttagcaaggctacgaagactacgttaa 287  
Db 141 AAAATTTTCTGAAAGTTTGAATGATTCCTAGAACTTTGGCTGAGAATGCTGGGTGAA 200  
Qy 288 tggaaagaacacctaactggtttatttgaggaaacttttaggggactcagaagaagctgg 347  
Db 201 TGCAATGGAGATCATATCTTCTCTGTATGACAGACATGCTGGAATCCCAAGTTGG 260  
Qy 348 agactggctctccacagtaactgtatgagtatgctgaagccttttgcagttccagaataaga 407  
Db 261 CATTGATTTGGAAGAGCGGTATGTAAGGATGTCTCAACCTTGAGCATTTGGGATCTCCA 320  
Qy 408 tgt 410  
Db 321 TGT 323

RESULT 2

LOCUS AQ769224 498 bp DNA GSS 28-JUL-1999

DEFINITION

Accession  
Version  
Keywords

Source

Organism

Reference

Authors

Title

Journal

Medline

Comment

HS\_3155\_B2.G11.T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3155 Col=22 Row=N, DNA sequence.  
AQ769224  
Version AQ769224.1 GI:5647340  
Keywords GSS:  
Source human.  
Organism Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 498)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3687  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end web server: http://www.htsc.washington.edu  
Plate: 3155 row: N column: 22  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 498.  
Location/Qualifiers

FEATURES

source

1. .498

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=3155 Col=22 Row=N"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 182 a 83 c 85 g 145 t 3 others

ORIGIN

Query Match 5.9%; Score 38.2; DB 13; Length 498;

Best Local Similarity 47.7%; Pred. No. 4.5;

Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 363 agtacgttatgagtatgtcgaagccttgctcagttccagaaatagatgttttcaggattgg 422

Db 289 ACTAGTTTGGGGATTATTCTTGATTCCTCTTCTGCTAGTAGAAATTTGAAGTACATG 230

Qy 423 ccgttgtaattatttaaaagtttgggtggcccaagcaattgctgaactatgacctaa 482

Db 229 CTATTCTACTTAATATTTATTTTGAACACACTCTCCCAAGTCTATCCATTATTCAGACCA 170

Qy 483 agagactaatttttacaattataaaagattttccgcctctatatgtatggatcac 542

Db 169 TGTTCGAATAACTCTCAATGTTTGTAGATATTTTGACTGTATATTTTCTGCTTAA 110

Qy 543 agatctctatcattcagagcttatggggcttactccaaaccagcaaacgataaa 597

Db 109 GAATTTTATATTATCTGTCTGAAATGGCCTCCAGTGAACACGATATCAAGAAA 55

RESULT 3

LOCUS AK013338/c

DEFINITION

Accession

Version

Keywords

Source

AK013338 1894 bp mRNA HTC 05-JUL-2001  
Mus musculus 10, 11 days embryo cDNA, RIKEN full-length enriched library, clone:2810453I12, full insert sequence.  
AK013338  
CAP trapper.  
Mus musculus (strain:C57BL/6J) 10, 11 days embryo cDNA to mRNA,  
CAP trapper.  
Mus musculus (strain:C57BL/6J) 10, 11 days embryo cDNA to mRNA,  
CAP trapper.



FEATURES

source

Location/Qualifiers

1. .1894

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/db\_xref="MGD:MGI:1905655"

/db\_xref="MGD:MGI:1920026"

/clone="2810453L12"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="10, 11 days embryo"

132. .1814

/note="putative"

/codon\_start=1

/protein\_id="BAB28799.1"

/db\_xref="GI:12850640"

/translation="MSQVLFQQLVPLLVKCKDRERRGSRVSIELQSLNSNPVHRKDL  
VIRLDDTPFFLYLVISEEDFQSLKLOQLGVDFLAPQKFDIDLLQCMQEHAKET  
PREFLLQSLLSLNSPVLNVVETNPFKHLHLSLKLLPGNDVDFIKKFLAGCLKAK  
KEEKLRSLSDDVTRLOLHITQETLSEKMQEIDLKRSWASHTASITNKHKSQELTAEK  
EKALOTVOCQOHEQKKELETLHORNHLQLSRLESELAANKELTERKYKGDSTVR  
ELKALAGVESELQAKQVLSRINCHLDTTECHKRKHINQKTVAVLQEIADKDK  
DLVLRTEAFDTIQEQVVALENGEKNOIQIGLKEATIKLSAEALLKANEIILKLOL  
DLNTLGMKLKLVNTYIQEKLAAEKEMLOKREKESQDAGFLRAKEQEVRLQLOL  
ETTVQKLEESKOLLNNEKLITWLKNELENQIVRKQDTLGTSATPHSTSNSTIRSLG  
SPNLNVDRNLNPSGCGYGPVSALTTFQNAFPHVVAKNKTSHPISGPKVAKC"

BASE COUNT 634 a 397 c 433 g 429 t 1 others

ORIGIN

Query Match 5.8%; Score 37.8; DB 12; Length 1894;

Best Local Similarity 60.0%; Pred. No. 6.9;

Matches 63; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 424 cgtggaattattaaagtttggttcgcccagcaagtaattgctctaactatgactctaaa 483

Db 1306 CCTGTGAACCTCTGTGATGATCTCATTTGCCCTTTAGAAGTGTCTGCTGATAATGATTTTATT 1247

QY 484 gagcgtaagttgttacaattataagagatttcgcgctctatat 528

Db 1246 GTAGCTCAAGTTTCCCACTTGTATCTGATTTTCTCACCATT 1202

RESULT 4

AZ9333264

LOCUS

DEFINITION

GM135b001.045\_J24R ISU Soybean BAC Library (pBeloBAC11 HindIII)

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Glycine max

446 bp DNA GSS 04-APR-2001

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

1 (bases 1 to 446)

Larson,K., Mudge,J., Cooper,A., Grivna,S., Denny,R., Penuela,S.,  
Danesh,D. and Young,N.D.

EAC End sequences from a soybean genomic library (UMN)

Unpublished (1999)

Contact: Young Nevin D

Department of Plant Pathology

University of Minnesota

495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul MN 55108, USA

Tel: 612 625 2225

Fax: 612 625 9728

Email: nevin@tc.umn.edu

Sequence on contig Gm.A685\_ctg\_a near unmapped duplicate of pA685.

For more information, see SoyBase at:  
[http://genome.cornell.edu/cgi-bin/WebAce/webace?db=soybase&class=Lo-  
cus](http://genome.cornell.edu/cgi-bin/WebAce/webace?db=soybase&class=Lo-<br/>cus). Please see as an authority for the mapping/naming: Cregan P.B.  
T. Jarvik, A.L. ush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N.







Db	221 T 221	
RESULT	7	
LOCUS	AA185888	
DEFINITION	AA185888 412 bp mRNA EST 17-FEB-1997 mt35c07.r1 Soares mouse 3NbMS Mus musculus CDNA clone IMAGE:623052	
ACCESSION	AA185888	
VERSION	AA185888.1 GI:1772287	
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 412) Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, P., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.	
TITLE	The WashU-HMMI Mouse EST Project	
JOURNAL	Unpublished (1996)	
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.	

```

email: mouseestewatson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:383876

Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 406.
Location/Qualifiers
1. .412
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:623052"
/clone_lib="Soares mouse 3NbMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"

/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAACTCAAGTGGAGCGCGCGCTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Ratima Bonaldo."

121 a 65 c 70 g 156 t

```

[illegible]







[illegible]

RESULT	12			
CNS0039G/C				
LOCUS	CNS0039G	1101 bp	DNA	GSS
				03-JUN-1999



D	B	741	RAGTAGRKWRWTKRWRKRRDTRWDADADDATDRDRRCDDGADAGKGKKTKGRKRRR	682
Q	Y	487	gctaatagttttcaacaattataaaaggagttttccgcgtctatataatgatgcatcacagat	546
D	B	681	DRA'TWDRTDAAWADAANWTDTDDDKDRDRKRKGARRRRRTTARAANDWMTWKAWDW	622
Q	Y	547	tctctacattcagagcttatggcgcttactccaaccagcaaacgataaaactcgcagt	606
D	B	621	AKDWKTRADRWDRWAADTWTDARKADWNDAKARWARRDRAARAADRSWTTKGKTTT	562
Q	Y	607	gattttactttccgaagaagttagcttagtgtaataattt	643
D	B	561	ATWTTWAARAANAANAATAATTATWTTTWTWTTT	525

RESULT	13
CNS0106X	
LOCUS	CNS0106X ll101 bp DNA GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BACN03K20 of DrosBAC library from drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL098595
VERSION	AL098595.1 GI:5610206
KEYWORDS	GSS.
SOURCE	fruit fly. Plasmid Drosophila melanogaster
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridioidea; Drosophilidae; Drosophila.
REFERENCE	
AUTHORS	(bases 1 to ll101)
TITLE	Genoscope.
FEATURES	Direct Submission

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES Location/Qualifiers

source	1..1101
	/organism="Drosophila melanogaster"
	/plasmid="pBelobAC11"
	/db_xref="taxon:7227"
	/clone_lib="DrosBAC"
	/clone="BACN03K20"
	/note="end : T7"

BASE COUNT 258 a 107 C 60 g 175 t 501 others

ORIGIN

Query Match 5.6%; Score 36.4; DB 13; Length 1101;  
Best Local Similarity 14.5%; Pred. No. 16;  
Matches 38; Conservative 125; Mismatches 99; Indels 0; Gaps 0;

Qy 188 atagtcagtcttccatgaaggcactgaagaacgaccttatctttatggaggtttct 247  
| | | | | : | | | | | : | | | | | : | | | | | :  
Db 693 AAATCATCKDARKAKAKAKATRKAKARAWDTATWTATAADAAADKGAKDKAKAKAD 752

Qy 248 taatgaacccttttagcaaggctcaagaagctacgtttaattgaaagaaaaccttgctt 307  
| | | | | : | | | | | : | | | | | : | | | | | :  
Db 753 ARDKRRWDAKKRKRRAKAKAADDDADAADAKADAKAADAADGGRGGKKR 812

Qy 308 ggattattggagaaactttaggggactcagaagaagctggagactggtctgcacagtlac 367  
| | | | | : | | | | | : | | | | | : | | | | | :  
Db 813 AKDRKKKKDKAWGDKKKAKDKAAAKADADAGDAKAKRRRAGDKOKAKADAKAKKA 872

Qy 368 gttatgatgtatcgaaagccttgtcagttccogaataatagatgtttcagggatggcgtg 427  
| | | | | : | | | | | : | | | | | : | | | | | :







Db 684 CCATGAATGACAAACCAACCAATTTCCATATGGTATTTTCTTAAAAAACTTCATAAGA 743  
Qy 267 ggctacgaagactacgtttaaattggaagaaacccctagcttggtttattggaggaaacttt 326  
Db 744 CACTATTGCAATGCATAAACATCAATAGGGAATGAAGTAGGGTTTATTTAATAAAATTT 803  
Qy 327 agggggactcagaaagctggagac 351  
Db 804 CAGGTCCCATATACAAATGGGAGTC 828

Search completed: February 7, 2002, 21:32:15  
Job time: 20649 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: February 7, 2002, 19:27:24 ; Search time 12230.8 Seconds  
(without alignments)  
1982.766 Million cell updates/sec

Title: US-09-391-606-3

Perfect score: 1470

Sequence: 1 atgtgtaactctattgtgc.....accatcatcatcatcat 1470

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.om.\*

20: em.or.\*

21: em.ov.\*

22: em.pat.\*

23: em.ph.\*

24: em.pl.\*

25: em.ro.\*

26: em.sts.\*

27: em.sy.\*

28: em.un.\*

29: em.vi.\*

30: em.htgo.hum.\*

31: em.htgo.inv.\*

32: em.htgo.rod.\*

33: em.htg.hum.\*

34: em.htg.inv.\*

35: em.htg.rod.\*

36: em.htg.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1460.4	99.3	2238	6	AX045137	AX045137 Sequence
2	1389.4	93.2	2148	1	CHT76KDA	L23921 Chlamydia p
3	1359.4	92.5	2156	6	AX045131	AX045131 Sequence
c	1359.4	92.5	10421	1	AE001654	AE001654 Chlamydia
5	1359.4	92.5	12548	1	AE002165	AE002165 Chlamydia
c	1359.4	92.5	300550	1	AP002547	AP002547 Chlamydia
7	1347	92.5	1456	6	AX045135	AX045135 Sequence
8	1155.4	78.6	1852	6	AX045133	AX045133 Sequence
9	146.8	10.0	150	6	ARI44067	ARI44067 Sequence
10	65	4.4	816	6	AX105043	AX105043 Sequence
11	65	4.4	1200	6	AX092220	AX092220 Sequence
12	65	4.4	1373	6	AX105044	AX105044 Sequence
13	65	4.4	1620	6	AX105036	AX105036 Sequence
c	65	4.4	1620	6	AX105037	AX105037 Sequence
14	64.2	4.4	6849	12	AY029367	AY029367 Synthetic
15	63	4.3	4357	6	AX072889	AX072889 Sequence
16	63	4.3	7108	6	AX001326	AX001326 Sequence
17	63	4.3	7108	6	E36262	E36262 Human semap
18	63	4.3	7475	6	AX001323	AX001323 Sequence
19	63	4.3	7475	6	E36259	E36259 Human semap
20	63	4.3	8192	6	AX001324	AX001324 Sequence
21	63	4.3	8192	6	E36260	E36260 Human semap
c	56.6	3.9	10954	1	AE001333	AE001333 Chlamydia
24	56.4	3.8	72	6	AX019231	AX019231 Sequence
c	54.4	3.7	7218	6	I66494	I66494 Sequence 14
25	53.6	3.6	1537	6	AX155814	AX155814 Sequence
26	51	3.5	5098	1	MCU51235	U51235 Mycoplasma
27	50.6	3.4	125	6	AR077332	AR077332 Sequence
28	50.6	3.4	2085	6	AR008322	AR008322 Sequence
29	50.6	3.4	2085	6	AR027306	AR027306 Sequence
30	50.6	3.4	2085	6	AR035771	AR035771 Sequence
31	50.6	3.4	2085	6	AR037631	AR037631 Sequence
32	50.6	3.4	2085	6	AR078816	AR078816 Sequence
33	50.6	3.4	2085	6	AR079063	AR079063 Sequence
34	50.6	3.4	2085	6	AR084296	AR084296 Sequence
35	50.6	3.4	2085	6	I16740	I16740 Sequence 1
36	50.6	3.4	2085	6	I70472	I70472 Sequence 1
37	50.6	3.4	2085	6	I92332	I92332 Sequence 1
38	50.6	3.4	2086	1	STRSURPROA	M74122 Streptococc
39	50.4	3.4	217	6	AR083872	AR083872 Sequence
40	50.4	3.4	807	9	HS225093	AJ225093 Homo sapi
41	50.4	3.4	819	9	HS225092	AJ225092 Homo sapi
c	50.2	3.4	14698	1	AE002357	AE002357 Chlamydia
43	50.2	3.4	14698	1	AR054408	AR054408 Sequence
44	49.8	3.4	913	6	AX080641	AX080641 Sequence
45	49.8	3.4	913	6	AX080641	AX080641 Sequence

## ALIGNMENTS

### RESULT 1

AX045137

LOCUS

DEFINITION

AX045137

ACCESSION

AX045137.1

VERSION

GI:11343736

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

PAT

24-NOV-2000

AX045137

Sequence 7 from Patent WO0066739.

AX045137

AX045137.1

GI:11343736

Chlamydia pneumoniae.

Chlamydia pneumoniae

Bacteria; Chlamydiales;

Chlamydiae; Chlamydia

1 (bases 1 to 2238)

Murdin, A.D., Oomen, R.P., Wang, J. and Dunn, P.

Chlamydia pneumoniae

thereof

Patent: WO 0066739-A 7 09-NOV-2000;

Aventis Pasteur Limited (CA)

Location/Qualifiers

1. 2238

/organism="Chlamydia pneumoniae"

/db\_xref="taxon:83558"



CDS	766..2238	/note="unnamed protein product"	
	/codon_start=1		
	/transl_table=1		
	/protein_id="CAC17330.1"		
	/db_xref="GI:11343737"		
	translation="MVNPIGPGPIDETPPADLSAQLSEASNAKSAFAORIGAE AKPKESKTSYVERWSILRSVNALMSLADKLGIASSNSSTSRSDVDSTTATPTP PPPTFDYKTAQAYDTIETSLADIQAALVSLQDAVINIKDTRATDETAIAEW ETKNADVKYGAQITELAKYSDNOAILDSGLKTSFDLLQALLOSVAANNKAAELL KEMQNDVPVKPTPAIAQSLVDQTDATATQIERKDNARIDAYFAGNAGAVENAKSN NSISNIDSKAQAATKTQIAEAKKFPDPIQEAEQMWIOAEKDLNKKPADGSDV PNPCTTGVGSKQGSSTISRVSMILLIDENETASILMSGFROWIMHFNTPNDSDAA OQELAAQARAKAGDSDSAAALADACKLEAALGRAGQOQIILNALGQIASAVVSA GVLPQLVIRARYQAYVEQKLISEDLNSVDHHHHH"		
BASE COUNT	689 a	459 c	497 g 593 t
ORIGIN			
Query Match	99.3%	Score 1460.4;	DB 6; Length 2238;
Best Local Similarity	99.6%	Pred. No. 0;	
Matches 1464;	Conservative 0;	Mismatches 6;	Indels 0; Gaps 0;
Qy	1	atggttaacccattggtccaggctcctatagcgaacacagacacacacccctcccgagat	60
Db	766	ATGGTTAATCTTATGGTCCAGGCTCTATAGCGAAACAGAGACGACACCTCCCGCAGAT	825
Qy	61	ctttctgcacagattgagggcagtcagcaataagagtgcggaagctcaagaata	120
Db	826	CTTTCTGCTCAAGATTGGAGCGAGTCGACCAATAAGATGGCGAAGCTCAAGAAATA	885
Qy	121	gcagtgcggaagctaaagcctaaagaatcgaacccgattctgtagagcgatgagcattc	180
Db	886	GCAGGTGCGGAAGCTAAGCCTTAAGAACTTAGACCGATTCTGTAGACGGATGGAGCATC	945
Qy	181	ttgcgtttcgcagtggaattctctatagcttgagctggcagataagctgggtattgcttctagt	240
Db	946	TTGCGTTCTGCAAGTAAATGCTCTCATAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT	1005
Qy	241	aacagctgctcttactagcagatctcagacgtggactcaacgacgacgacgacccct	300
Db	1006	AACAGCTGCTCTTCTACTAGCAGATCTGCAGCGTGGACTCAACGACAGCGACCCGACCT	1065
Qy	301	acgcctctcccccacgctctgatattataagactcaagcgcaaacagcagcttacgatact	360
Db	1066	ACGCTCTCTCCACCCAGGTTTGATGATTATAGACTCAAGCGCAACAGCTTACGATACT	1125
Qy	361	attcttaacctcaactcactagctgacatacaggtctgttggtgagcctccaggatgct	420
Db	1126	ATCTTTACCTCAACTCACTAGCTAGCTGACATACAGGCTGCTTTGGTGGCTCCAGGATGCT	1185
Qy	421	gtcactaatataagagatcacgctgctactgatgaggaacccgcaatcgctgcggagtg	480
Db	1186	GTCACTAATATTAAGGATACACCGGCTACTGATGAGGAACCGCAATCGCTCGGAGTGG	1245
Qy	481	gaaactaagaatgccgatgcaattaaagtggcgcgcaaaattacagaattagcgaatat	540
Db	1246	GAACTAAGAATGCCGATGCAGTTAAAGTTGGCGCGCAAAATTACAGAATTACGGAATAT	1305
Qy	541	gtctgggataacacagcgattcttgaactcttaggttaactgaactccttcgaaccttta	600
Db	1306	GCTTCGGATAACCAAGCGATTCTTGACTCTTTAGTAAACTGACTTCCTTCGACCTCTTA	1365
Qy	601	cagactgctcttctcaatctagcaacaataacagcagctgagcttctttaaagag	660
Db	1366	CAGGCTGCTCTTCTCCAACTCTGTAGCAACAATAATACAAAGCAGCTGAGCTTCTTAAGAG	1425
Qy	661	atgcaagataacccagtagtccacgggaacgctcgcaattgctcaattcttagttgat	720
Db	1426	ATCAAGATAACCCAGTAGTCCCGGGAACCGCTGCAATTCCTCAATCTTTAGTTGAT	1485
Qy	721	cagacagatgctacagcgacacagatagaaagatggaattgcgattgggagtcata	780

RESULT	2
CHT76KDA	2148 bp
LOCUS	Chlamydia pneumoniae 76 kDa protein gene, complete cds.
DEFINITION	Chlamydia pneumoniae 76 kDa protein gene, complete cds.
ACCESSION	L23921
VERSION	L23921.1 GI:435961
KEYWORDS	76 kDa protein.
SOURCE	Chlamydia pneumoniae DNA.
ORGANISM	Chlamydia pneumoniae
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
AUTHORS	1. (bases 1 to 2148)
TITLE	Perez-Melgosa, M., Kuo, C.-C. and Campbell, L.
	Isolation and characterization of a gene encoding a Chlamydia
	pneumoniae 76-kilodalton protein containing a species-specific
	epitope. Immun. 62, 880-886 (1994)
JOURNAL	Infect. Immun. 62, 880-886 (1994)
MEDLINE	94156481
FEATURES	Location/Qualifiers
	1..2148
	/organism="Chlamydia pneumoniae"
	/db_xref="taxon:83558"
	1..2148
CDS	



```

/codon_start=1
/transl_table=11
/product="76 kDa protein"
/protein_id="AA23117.1"
/db_xref="GI:435962"
/translation="MTKKHYAWVVEGILNRLPKQFFVKCSVVDWNTFVPSSTSTTEKA
ATNAMYKYCYWQWLVGHKSVQVINGQKPLLYLGFALMPLAKATKTLTGKLENLA
WFIGTGLKQKAGDWSATVVEIDVSGIGRNLKLFKFAOIAIANYDP
KEANFTNYKFSALYMGITDSLSFRAYGASKPANDKLGSDTFKFDLGLIISAGF
KSNENKIPKRLALIISSEFFIYFNKTKRELLFFELLVNPISGPNIDETERTPPAD
LSAGLESAANKSAEORIAEAKRESKIDSVERWSILRSVAVNLMSLADKLGA
SSNSSTSRADVDSTTATPPPTFFDDYKTOAQATYDTIFETSLADIAALVS
LODAVNTKTDAADEETAIAAEWETKNADAVKVGQAQITELAKYASNDQALIDSLGKL
TSFDLQAALQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVDOTDADATQIEKD
GNARDYFACONASGAVENAKSNISNIDSAKAAIATAKTOIAEAKQKFPDPSILO
EAEQWIOAERDLKNIKPADGSDVPNGPTTVGSKQOQSSIGSIRVSLMLDDAENETA
SILMSGQRMTHMENTENPDQSAQAQELAAARAKAAGDDSAALADAKALEAAL
GKAGQQQILNALQIAASAAVSGVLPLOQVL"
BASE COUNT      659 a 435 c 480 g 574 t
ORIGIN

```

```

Query Match      93.2%; Score 1369.4; DB 1; Length 2148;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1373; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2  tggtaaatcctattggtccagctctatagacgaacagcagcaccctcccgagatc 61
DB 770 TGGTTAATCTATTGGTCCAGGTCCTATAGACGAACAGCACACCTCCCGCAGATC 829
QY 62  ttctgctcaaggattggaggcgagtgcaagcaataaagagtcggaagctcaaaagatg 121
DB 830 TTTCTGCTCAAGGATTTGGAGCGAGTGCAGCAAAATAGAGTGCAGGAAAGCTCAAGAATAG 889
QY 122 caggcggaagctaagcctaagaatctaagacgagcttcttagagcgatgagcagcatc 181
DB 890 CAGGTGCGGAAGCTAAGCCCTAAGAAATCTTAAGACCGATTCTGTAGCGGATGAGCATCT 949
QY 182 tgcgttctcagtgaaatgctctcagctgagtcgagcaataagctggtgattgctctagta 241
DB 950 TCGGTTCTGCACTGAATGCTCTCATGAGTCTGCGAGATAAGCTGGGTATTGCTTCTAGTA 1009
QY 242 acagctcgtcttctactagcagatctgacagcgtggaactcaacagcagcagcagcacta 301
DB 1010 ACAGCTCGCTTCTTACTAGCAGATCTGCAGAGCTGAGCTCAACAGCAGCGACCGACCTA 1069
QY 302 cgcctcctccaccagctcgtgattataagacacacgcaacagcagcttacgatacta 361
DB 1070 CGCCTCCTCCACCCAGTTTGGATTATTAAGACTCAAGCGCAACAGCTTACGATACCTA 1129
QY 362 tctttactcaacatcactagctgacatagcagctgcttggtagcctccaggtgctg 421
DB 1130 TCTTTACCTCAACATCACTAGCTGACATACAGGCTGCTTTGGTAGGCTCCAGGATGCTG 1189
QY 422 tcaactaataaagatacagcgctactgtagtgaagaaaccgaatcgctcgaggtggg 481
DB 1190 TCACAAATAAAGGATACAGCGGCTACTGATGAGGAAACCCCAATCGCTGCGGAGTGGG 1249
QY 482 aaactaagaatccgatgcaataaagttggcgcaataattacagaattagcgaaatag 541
DB 1250 AAACTAAGATCCCGATGAGTTAAGTTGGCGCCCAATTAACAAATTAGGCAAAATAG 1309
QY 542 cttcggataaccaagcagattctgactctttaggttaaaactgacttcttcgacctcttac 601
DB 1310 CTTTCGGATAACCAAGCGATTCTTACTCTTTAGGTAAACTGACTTCTCTCGACCTCTTAC 1369
QY 602 agactgctcttctcactctgtagcaaacataaacaagcagctgagcttcttaagagaga 661
DB 1370 AGGCTGCTCTTCTCAATCTGTAGCAAAACATAAACAAGCAGCTGAGCTTCTTTAAAGAGA 1429
QY 662 tgcaagataaccagctagtcctccagggaaaacgctgcaattgctcaattcttagttgac 721
DB 1430 TCCAAGATPACCCAGTAGTCCCGGAAACGCCCTGCAATCTCTCAATCTTTAGTTGATC 1489

```

```

QY 722  eqacagtgctacagcagcacagatagagaagatggaatcgattggggatgcatt 781
DB 1490 ACACAGATGCTACAGCGACACAGATAGAGAAAGATGGAATGCGATTAGGGATGATATT 1549
QY 782  ttgcaggacagaaacgctagtgagctgtagaaaatgctaaaatcttaataacagataaagca 841
DB 1550 TTGCAGGACAGAAACGCTAGTGCAGCTGTAGAAAATGCTAAATCTAATACACATATAAGCA 1609
QY 842  acatagattcaactaaagcagcaatcgctactgcttaagacacaaatagctgaagctcaga 901
DB 1610 ACATAGATTCACTAAAGCAGCAATCGCTACTGCTTAAGACACAAATAGCTGAAGCTCAGA 1669
QY 902  aaaaagctcccgactctcaattcttcaagaagcgaacaaatgtaatacacagctcaga 961
DB 1670 AAAAGTTCCCGCACTCTCCAAATCTTCAAGAACGCGAACAATGGTAATACAGGCTGAGA 1729
QY 962  agagattctaaaaataatcaaacctcgagatggttctgtagtttcccaatcccaagaactacag 1021
DB 1730 AAGATCTTAAAAATATCAAACTCGCAGATGGTTCTGATGTTCCAAATCCAGGAACATACAG 1789
QY 1022 ttggagggtccaaagcaacaaagaaagttagtattggttagtattcgtgtttccatcgtgttag 1081
DB 1790 TTGGAGGCTCCAAAGCAACAAAGAAAGTAGTATTGGTAGTATTGCTGTTTCCATGCTGTAG 1849
QY 1082 atgatgctgaaaatgagaccgcttccatttgcatttgcgtggttctgagatgattcaca 1141
DB 1850 ATGATGCTGAAAATGAGACCGCTTCCATTTTGTATGCTGCGGTTTGTGTCAGATGATTACAG 1909
QY 1142 tgttcaatcaggaataatcctgattctcaagctgcccacacagagctcgagcacaagctca 1201
DB 1910 TCTTCAATACGAAAAATCTGATTCTCAAGCTGCCAACAGGAGCTCGCAGCACAAGCTA 1969
QY 1202 gagagcgaagaaacccgctggagatgacagctgctgctgagcgcgtgagatgctcagaaag 1261
DB 1970 GAGCAGCGAAAGCCCGCTGGAGATGACAGTGCCTGCTGCAAGCGTGCAGATGCTCAAGAA 2029
QY 1262 ctttagaagcgcctctagtaaagctggcgaacacacagggcactactcaatgctttgggag 1321
DB 2030 CTTTAGAGCGGCTCTAGGTAAGCTGGGCAACACACAGGCGCATACTCAATGCTTTAGGAC 2089
QY 1322 agatcgcttctgctgctgctgtagcagcaggagtcctcccgctgagcagaagttctatg 1380
DB 2090 ACATCGCTTCTGCTGCTGTTGTGAGCGCAGGAGTCTCCCGCTGCAGCAAGTCTCTATAG 2148

RESULT 3
AX045131 LOCUS
DEFINITION Sequence 1 from Patent WO0066739.
ACCESSION AX045131
VERSION AX045131.1
KEYWORDS GI:11343730
SOURCE Chlamydomophila pneumoniae.
ORGANISM Chlamydomophila pneumoniae
REFERENCE 1 (bases 1 to 2156)
AUTHORS Mordin, A. D., Oomen, R. P., Wang, J. and Dunn, P.
TITLE i(chlamydia) antigens and corresponding dna fragments and uses
JOURNAL thereof
PATENT: WO 0066739-A 1 09-NOV-2000;
Aventis Pasteur Limited (CA)
Location/Qualifiers
1. 2156
/organism="Chlamydomophila pneumoniae"
/db_xref="taxon:83558"
101..2056
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC17329.1"
/db_xref="GI:11343731"
/translation="MVNPIGPGPIDETERTPPADLSAOGLEASANKSAEORIAAG

```



AKPKESKTDVSERWSILRSVAVNALMSLADKLGIASSNSSSSRSADVDSTTATPTP  
 PPTDDYKQAOAYDTIFTSIADLOAALVSDAVTNKIDTDEETAIAAEW  
 ETKNADAKVGAOITELAKVASDNOAILDSGLKTSFDLLQALLOSIVANNKAAELL  
 KEMQDNVPVGGKTPAIAQSLVDOTDATAIOIEKDNRAIRDYFAGQNASCAVENAKSN  
 NSISNIDSAKAAATATAKTQAEQKFPDPILOEAEQVWIOAEKDLNLIKPADGSDV  
 NPNFTVGGSKQGSSTIGSRVSMLLDAENETASILMSGFROMHMFNTENPDSQAA  
 QOELAAQARAAGADSDSAALADAKLEAALGRAGQOQGLNALGQITASAAYVSA  
 GVPPAAASSTGSSVKOLYKTSKSTGSDYKTOISAGYDAYKSIDYAGRNADATROVI  
 NNYSPTALTRSVPRAEPKPTDOALARVIGNSRSLGVDYSOVSALQSYMOLIQ  
 SNPANNNEIRKOLITSVTPKPGPIYVOLSNDSTQKFIKLESLEFAESRSTAAEIK  
 ALSFETNSLFIQOVLNIGSLYSYLIQ\*

BASE COUNT 672 a 461 c 471 g 552 t  
 ORIGIN

Query Match 92.5%; Score 1359.4; DB 6; Length 2156;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 1 atgggttaacatctattggtccaggtctctatagacaaacagacacacacccctcccgagat 60  
 Db 101 ATGGTAAATCCATTGTTCCAGGTCTCTATAGACGAAACAGACGACACCTCCCGCAGAT 160  
 Qy 61 cttctctctcaaggattgagcgagtgagcagaaataagagtgoggaagctccaagaata 120  
 Db 161 CTTTCTGCTCAAGGATTGGAGCGAGTGCAGCAATAAAGAGTGCAGAAAGCTCAAGAATA 220  
 Qy 121 gcaggtggaagctaaagcctaaagaaatctaaagccgattctgtagacgagtgagcattc 180  
 Db 221 GCAGGTGGGAAAGCTAAGCCTAAAGAACTAAGACCGATTCTGTAGCGGATGGAGCATC 280  
 Qy 181 ttgcgtctcgagtgaatgctctcagtgatctggcagataagctgggtattgcttctagt 240  
 Db 281 TTGCGTCTGCAGTGAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT 340  
 Qy 241 aacagctogtcttctactagcagatctcgcagacgctgagcactcaacacgacgcacacct 300  
 Db 341 AACAGCTGCTCTTCTACTAGCAGATCTGCAGAGTCTGCAGAGTGGACTCAACGACAGCGACCGACCT 400  
 Qy 301 agcctctccaccacgctctgattattataagactcagcgaacagcattcagatact 360  
 Db 401 AGCGCTCCCTCCACCCACGCTTGTGATTATTAAGACTCAAGCGCAACAGCTTACGATACT 460  
 Qy 361 attcttaacctcaactcactagctgacatacaggtgctttgtgagcctccagatgct 420  
 Db 461 ATCTTTACCTCAACATCACTACTACTGACATACAGGTGCTTTGGTGAGCCTCCAGATGCT 520  
 Qy 421 gtcactaatataaaggatacagcggctactgtaggaaacccgcaatcgctgcggagtg 480  
 Db 521 GTCACATAATAAAGGATACAGCGGCTACTGATGAGGAACCGCAATCGCTCGGAGTGG 580  
 Qy 481 gaaactaagaatccgagatcaataaagttggcgcgaataattacagaattacgaaatat 540  
 Db 581 GAAACTAAGAATCCCGATGCAGTAAAGTTGGCGGCAATATACAGAATTACGGAATAT 640  
 Qy 541 gcttcggaataacaaacgagcttctgactcttaggttaactgactctcttcgaccttta 600  
 Db 641 GCTTCGGATAACACAGCGATTCTTGACTCTTTAGTAAACTGACTTCCCTTCGACCTCTTA 700  
 Qy 601 cagactgctctctccaatctctagcaaaataacaaagcagctgagcttctctaaagag 660  
 Db 701 CAGGCTGCTCTTCTCCAAATCTAGCAAAATAAACAAGCAGCTGAGCTTCTTAAAGAG 760  
 Qy 661 atgcaagataaaccagtagtccagggaaacgctcgaattgctcaatctttagttgat 720  
 Db 761 ATGCAAGATAAACCAGTAGTCCAGGGAAACGCCCTGCAATTCGTCATCTTTAGTTGAT 820  
 Qy 721 cagacagatgctacagcgacagatagagaaagtggaatgcgattggggatgcataat 780  
 Db 821 CAGACAGATGCTACACGACACAGATAGAGAAAGATGGAATGCGATTAGGGATGCATAT 880  
 Qy 781 ttgcggagcagaacgctagtgagctgtagaaaaatgctaaatcttaataacaggtataagc 840

RESULT 4  
 AE001654/c  
 LOCUS AE001654 10421 bp DNA BCT 01-DEC-2000  
 DEFINITION Chlamydia pneumoniae section 70 of 103 of the complete genome.  
 ACCESSION AE001654 AE001363  
 VERSION AE001654.1 GI:4377031  
 KEYWORDS  
 SOURCE Chlamydia pneumoniae CWL029.  
 ORGANISM Chlamydia pneumoniae CWL029  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 REFERENCE 1 (bases 1 to 10421)  
 AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hymen, R.W., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
 MEDLINE 99206606  
 PUBMED 10192388  
 REFERENCE 2 (bases 1 to 10421)  
 AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
 FEATURES  
 source Location/Qualifiers  
 1. .10421  
 /organism="Chlamydia pneumoniae CWL029"  
 /strain="CWL029"  
 /db\_xref="taxon:115713"  
 133..2751  
 /gene="CpH0727"



CDS

```

133. .2751
/gene="CpN0727"
/codon_start=1
/transl_table=11
/product="CT619 hypothetical protein"
/protein_id="AADI8866.1"
/db_xref="GI:4377032"
/transl_table="MSTFSIQNRLLTIGSGESTRIIKLDHKYSGFDPSPVAINLEELN
SGIVARLALNLOSENTNVAALLNPNTLEPTTSTWTDYKHSRPOASPRAPSTQPT
DIVSAALALVLDGLAELVASVTEIDGALSTISTVRQLMASYLGLTILTLAEQER
VFSSTVPSKLEHVKQKAAEQIAQKEIKAVLEAKGVSTEEAILIKFDPDIY
AADKFKEIEPLHYRAKVGAPIQOENNAIQLLTPPAITPDNVNVEVGMTLSTI
LQAIIDDAIKQAPALGQDEIITLQTLVPLVDKTTPTKAEDFLIYTATQLPNTASLK
YLDRQAEYRGKTKVYQNSIQNLSVETKRVVNNRSMLETLQSMFOAOAOCFVWTIS
QANALNTAITNKYISAVLTTSMEMYGGLLCISYMYERLADDEKALFDKSVNEXYLPJHI
VCGSWNGWTKAQAQELAEYSICTAVTSQDQIKAYLQTRGRNEFRATRRFFINIG
QMYQFANETVPGNCITANGAIQIDLGFG IREAMNVTGVEADYVSNQRLNNEFNTA
ATAHLVQLQIAELQKADDDPGKASFTEKFAVAANWITSESLGDLISMLNSQ
LPQOEAFLKPIIEEINFNNAALNSLLQITNEFSTTSVYSSLSYLQVSKTQGNLF
AGDYETLLAAAREVYIRDTARCKQAINLVNGLLOKINSLPGATSAQKQEMLNATT
YYQYSLVTNLQTLVLESLLAGLKWTLQTTSNKYDKSVKIESFDDWIPTLALESF
LTCGFPNISATGGLGPLEFTQVQSDQYTSQGTQQLNLQNQMTTIQEWTLVSTSMQ
VLNGLISQAGAIYSN"
complement(3078. .5033)
/gene="CpN0728"
complement(3078. .5033)
/gene="CpN0728"
/codon_start=1
/transl_table=11
/product="CHLPN 76 kDa Homolog_1 (CT622)"
/protein_id="AADI8867.1"
/db_xref="GI:4377033"
/transl_table="MVNPIGPGPIDERTPPADLSAQGLEASAANKSAEAQTAGAE
AKPKESKTDSEVERSLRSNALMSLADKIGIASNSSSTSRSDVDSPTATAPTP
PPPTFDYKQAOAYDTIFETSLADIQALVSLQDAVNIKTATIDETETAIAEW
ETKNADAVKVGQATIELAKYASDNQAILDSKGLTSFDLQALQLQSVANNKAAELL
KEMQNDVPVQKPTAIAQSLVDQDATQIEKQGNALRDAYFAQNAQSAVENAKSN
NSIISIDSAKAIATARTQIAEQKFPDPSILQEAQMYIAEKDLKNIKPADGSDV
PNPTTGVGSKQSGSISIRVSMLLDAENETASILMSGFRQIMHFNTEPNDSQAA
QOELAAQARAQAADSDSAALADAQAALFAALGKAGQOQILNALGOIASAVVSA
GVPPAAASSIGSSVKQLYKTSKTSQSDYKTOISAGYDAYKSIDNAYGRANDATRDVI
NNVSTPALTRSPVPRTEARGPEKTDQALARVISGNSRTLDGVYSOVSALQSVQIIQ
SNPOANNEEIRQKLTSAVTRPPQGYPIVQLSNDSTOKFIKLESLSFAEGSRTAAEIK
ALSFTNSLFIQQVLVNMISLSYGLQ"
complement(5145. .6413)
/gene="CpN0729"
complement(5145. .6413)
/gene="CpN0729"
/codon_start=1
/transl_table=11
/product="CHLPN 76 kDa Homolog_2 (CT623)"
/protein_id="AADI8868.1"
/db_xref="GI:4377034"
/transl_table="MKQVYQMLASVVLALTSYAEPLSEQVKVSKHTYTTILDEVK
DYLKRGFEVTRKQDGLVRITAGDVRARWLFREDIKNPSDKDKNPLPNVNRSEFYL
YIDTRAKRNLUSSKMNWTATAGGNTAAGVDINRAFLGRFYKPNPRTDFEFMEIGRS
GLGDLFESEQVFNDFGLHIYTRFELSXDYPQYIVHGGPFVYNNMTKKHYAVMEIGI
LNRLPKQFVKQSVVDNTEFVPSSTTEKAAATNAMKYKYCVQMQLVKGKSOQVPIWG
QKPKLYLYGAPLMNPLAKATKTLNKENLAWFTGGTLGLLRKAGDWSATVRYEYVEA
LSVPEIDVSGIGRGNLLKFWFAQIAAANYDPEKEANGFTNYKGFSAALYMGITDSLFR
AYGAYSXPANDKLGSDFTFRKFDLGIISAF"
complement(6516. .8159)
/gene="mvn"
complement(6516. .8159)
/gene="mvn"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AADI8869.1"
/db_xref="GI:4377035"
/transl_table="MSRKDNVSLARSIFENILSGTFCSRITGIFREIATAMATYFGADPI
VAAPWLGRTVFFLRLKILGLLLEQAFIPHFEFLRAQSLDRAAFRRFRFLKSGTII

```

gene

CDS

```

133. .2751
/gene="CpN0731"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8870.1"
/db_xref="GI:4377036"
/transl_table="MSFLLIKFPSPPLKQDYLFDISPYTSSSEITIGSYFKLNKASL
QSSTLRLRSISIIIS"
/gene="nfo"
complement(9654. .10283)
/gene="nfo"
complement(9654. .10283)
/gene="rs4"
/codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AADI8871.1"
/db_xref="GI:4377037"
/transl_table="MKVLPPPSIPLGAHTSTAGLKNAIYEGRDIGASTVQIIFTANO
QCMQRRALKKEVIEDFAALKETDLSYIMSHAGYLINPGADPDVILKRSRIGIYOEIL
DCTIGLTSFVNFHFGAALKSKSKEDMKNKIVSSQSAPLFDSPPLVLLTETTAGOQT
LIGSNFEELGVLVONLKNQIPIGVCDTCHIFAAGYDITSPQGHEDVLNEDEYVGLS
YLRAFHLNDSMFLGKANKDRHAPLGEYIGKESFKFLMDERTKRKPKYLETGPGPEN
WOKIGELLKFSKNRDS"
complement(9654. .10283)
/gene="rs4"
complement(9654. .10283)
/gene="rs4"
/codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AADI8872.1"
/db_xref="GI:4377038"
/transl_table="MARYCGPKNVARREFGANIFGRSNPLLLKKPHPPGOGHGMORKKK
SDYGLQLEEKQKAKACYMIMEKOLVKAKEVIRKQGNVAQMFLEPERECRLDNVYRM
GFARTIFAAQQLVAHGHLVNRVRRRPFPRGMQISLAKRSKRLOSQSDALESKD
ESSLPYSILDKTGFKGELLVSPEDQDIEAQLPLPINISVVCEFLSHRT"

```

BASE COUNT

ORIGIN

3278 a 2158 c 1983 g 3002 t

Query Match 92.5%; Score 1359.4; DB 1; Length 10421;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Oy 1 atgggttaatcctattggtccaggctcctataacgaaacagacacacacctcccgacat 60  
 |||||  
 Db 5033 ATGGTTAATCCTATTGGTCCAGGTCCTATAGCGAAGACAGACACCTCCCCAGAT 4974  
 Oy 61 ctttctgctcaaggattggaggcgagtgacgaaataagagtcgcggaagctcaagaata 120  
 |||||  
 Db 4973 CTTTCTGCTCAAGGATTGGAGCGGAGTGCAGCAATAAGAGTGCAGGAGTCAAGAATA 4914  
 Oy 121 gcaggtgcggaagctaagccctaaagaaacttaagaccgattctctagagcgatggagcatc 180  
 |||||  
 Db 4913 GCAGGTGCGGAAGCTAAGCCTAAAGAAATCTAAGACCGGATTCGTAGAGCGATGGAGCATC 4854  
 Oy 181 ttgcgttctgcagtgtaagtctctcatgagctctggcagataagctgggtattgctctctagt 240  
 |||||  
 Db 4853 TTGCGTCTCGAGTGAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTTAGT 4794



QY 241 aacagctgcttctactagacagatctgcagacgtgagctgaacacagcagccacact 300  
|||||  
Db 4793 AACAGCTGCTCTTCTACTAGCAGATCTGCAGACGTGGACTCAACGACAGCGACCGACCT 4734  
QY 301 acgctctctcaaccacagctctgatattataagactcgaagcgaacagcttactagact 360  
|||||  
Db 4733 AGCCTCTCTCCACCACGCTTGATGATTATAGACTCAAGCGCAACAGCTTACGATACT 4674  
QY 361 atctttacctcaactcactagctgacatacagagctgcttggtagcctccagatgct 420  
|||||  
Db 4673 ATCTTTACTCAACTCACTAGCTACATACAGGCTGCTTTGGTAGCCTCCAGATGCT 4614  
QY 421 gtaactataataagatacagcgctactgatgaggaacgcgaactcgcgcgagtgag 480  
|||||  
Db 4613 CTCACATAATAAGGATACAGCGCTACTGATGAGAAACCGCAATCGCTGCGGAGTGG 4554  
QY 481 gaaactaagaatgcgagtcgaattaaagttggcgcaattacagaattagcgaatat 540  
|||||  
Db 4553 GAAACTAAGAATGCCGATGCGATTAAAGTTGGCGCGCAATTACAGAATTAGCGAAATAT 4494  
QY 541 gcttcggataaaccagcattcttgaactctttagttaaactgaactccttcgaccttta 600  
|||||  
Db 4493 GCTTCGGATAACCAACGATCTTGCATCTTTAGGTAACTGACTTCCTTCGACCTCTTA 4434  
QY 601 cagactgctcttctccaatctgtagcaacaataaacaagcagctgagcttctttaaagag 660  
|||||  
Db 4433 CAGGCTGCTCTTCTCCAACTCTGTAGCAACAATAACAAGCAGCTGAGCTTCTTAAAGAG 4374  
QY 661 atgcaagataacccagtagtccccagggaaaaagcgcgaattgctcaatctttagttgat 720  
|||||  
Db 4373 ATGCAAGATAACCCAGTAGTCCCGAGGAAACGCTGCAATTCCTCAATCTTTAGTTGAT 4314  
QY 721 caacagatgctacagcagacagatagaaaagatgaaatgcgattggagatgcacat 780  
|||||  
Db 4313 CAGACAGATGCTACAGCCACACAGATAGAAAAGATGGAATGCCATAGGATGCCATAT 4254  
QY 781 ttgcagagacagaacgctagtgagctgtgaaaaatgctaaatcctaataacagtaagc 840  
|||||  
Db 4253 TTTGCAGGACAGAACGCTAGTGGAGCTGTGAAAAATGCTAAATCTAATAACAGTATAAGC 4194  
QY 841 aacatagattcagctaaagcagcaatcgtactgctgaagacacaaatagctgaagctcag 900  
|||||  
Db 4193 AACATAGATTACGCTAAAGCAGCAATCGCTACTGCTAGACACAAATAGCTGAAGCTCAG 4134  
QY 901 aaaaagttcccgagctctccaatcttcaagaagcggacaaatgtaatacagagctgag 960  
|||||  
Db 4133 AAAAAGTCCCGAGCTCTCCATCTTCAAGAGCGGGAACAAATGGTAATACAGCTGAG 4074  
QY 961 aaagatcttaaaaaatatcaaacctgcagatggttctgattgttcccaaatccaggaactaca 1020  
|||||  
Db 4073 AAAGATCTTAAAAATATCAAACTGCAGATGGTTCTGATGTTCCTCAATCCAGGAACTACA 4014  
QY 1021 gttgagagctccaagcaagaagtagattgtagtattctggttttccatgctgtta 1080  
|||||  
Db 4013 GTTGAGGCTCCAACCAACAGGAAGTAGTATTGTTAGTATTTCGTGGTTTCATGCTGTTA 3954  
QY 1081 gatgagctgaaaaatgagaccgctcccatcttggattgctggtgttcgcagatgattcac 1140  
|||||  
Db 3953 GATGATGCTGAAATGACAGCCGCTTCCATTTTGATGCTCGGTTTCGTCAGATGATTCAC 3894  
QY 1141 atgttcaatacggaaaaatctctgattctcaagctgccccacagggagctgcgagcacaagct 1200  
|||||  
Db 3893 ATGTTCAATACGAAAAATCCTGATTCTCAAGCTGCCCCAACAGGAGCTCGCAGCACAACT 3834  
QY 1201 agagcagcgaagcgcgtggagatgacagtgctgctgcagcgcgtgcagatgctcagaaa 1260  
|||||  
Db 3833 AGAGCAGCGAAGCGCGTGGAGATGACAGTGTCTGTCAGCGCTGGCAGATGCTCAGAAA 3774  
QY 1261 gctttagaagcgctcttagttaaagctgggcaacaacagggcactactcaatgctttggga 1320  
|||||  
Db 3773 GCTTTAGAAGCGGCTCTAGGTAAAGCTGGGCAACACAGAGGCGCACTCAATGCTTTAGGA 3714  
QY 1321 cagatcgcttctgctgtgtgtgtgagcgcaggag-tctcccgctgcagcaggttctatg 1379

Db 3713 CAGATCGTCTTCTGCTGCTTTGTGAGCGCAGGAGTTCCTCCGCTGCAGCAAGTCTCTATA 3654  
QY 1380 g 1380  
Db 3653 G 3653  
RESULT 5  
AE002165 12548 bp DNA BCT 30-MAY-2000  
LOCUS Chlamydomophila pneumoniae AR39, section 2 of 94 of the complete  
DEFINITION genome.  
ACCESSION AE002165 AE002161  
VERSION AE002165.1 GI:7188948  
KEYWORDS  
SOURCE Chlamydomophila pneumoniae AR39.  
ORGANISM Chlamydomophila pneumoniae AR39  
REFERENCE 1 (bases 1 to 12548)  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
TITLE Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39  
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
MEDLINE 20150255  
PUBMED 10684935  
REFERENCE 2 (bases 1 to 12548)  
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source  
Location/Qualifiers  
1. .12548  
/organism="Chlamydomophila pneumoniae AR39"  
/strain="AR39"  
/db\_xref="taxon:115711"  
/note="synonym: Chlamydia pneumoniae AR39"  
297. 965  
/gene="CP0011"  
297. .965  
/gene="CP0011"  
/note="similar to SP:P31218 GB:X71492 PID:296947 GB:U00096  
PID:1736770; identified by sequence similarity; putative"  
/codon\_start=1  
/product="uridine kinase"  
/protein\_id="AAF37907.1"  
/db\_xref="GI:7188950"  
/transl\_table=11  
gene  
/product="conserved hypothetical protein"  
/protein\_id="AAF37908.1"  
/db\_xref="GI:7188951"  
CDS  
complement(1047. .2021)  
/gene="CP0012"  
complement(1047. .2021)  
/gene="CP0012"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
product="conserved hypothetical protein"  
/protein\_id="AAF37908.1"  
/db\_xref="GI:7188951"  
/translation="MEKKYALAYYYITRYDNPHEIALHKFLDLDVSCRIVISEQ  
VNSK"



```

GINGOFSGYEPAHEALYMQWLKERPNFNSKIKFIHIIKENIIFRITVYRKELAAAGCE
VDLSKQAHISQEWHEKILQGNRCLILIDVRNNYEWKIGHFDNATLPDIOTRERPEYA
EKLAEQCPETTPVMYCTGIRCELYSPVLEKEGKEFYOLDGCVIAYGOQVGTGW
LGLKLFVEDRLAIPIDESDPVAPIAECCHQCTPSDAYNCANTDCHALFCCDECIH
QHQCCEGECQSQRVRKFDSSRGNKPFRAHLCEISENSEASCCLI'
2238..2867
/gene="CP0013"
CDS
2238..2867
/note="similar to SP:P81288; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S4"
/protein_id="AAF37909.1"
/db_xref="GI:7188952"
/translation="MARYCGPKNRVARREGANIGRSRNPILKKPHPPQGHGMORKKK
SDYGLQLEKOKKACGYMIMEKQLVKAPEVIHKQGNVAQMFLERFECRLDNMVRM
GFAKTIFAAQQLVAHGHLTVNGRVRDRSFFLRPMQISLSEKSKRQSVKDALESKD
ESSLPYSISLDTGTFKGELLYSPQDQTEAQLPLPINISVVVCEFLSHT"
complement(2986..3867)
/gene="CP0014"
CDS
complement(2986..3867)
/note="CP0014"
/note="similar to SP:P12638 GB:M22591 PID:146954
PID:405898 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="endonuclease IV"
/protein_id="AAF37910.1"
/db_xref="GI:7188953"
/translation="MKVLPPPSIPLLGAHTSTAGLKNAIYEGRDIGASTVOITFANO
ROWORALKREVIDEFAALKETDLSYIMSHAGVILNPGADPDVILEKSRIGIYQEIL
DCITLISFVNEHPCALKSKEDCMKNIVSFSQSAPLEDSSPLVLLVETTAGQGT
LIGNFEELGVLONKQIPIGVCVDCHIFAAGYDITSPQGWEDVNEDEIVEGLS
LYRAFHLNMFPLGANKDRHAPGEGYIGKESFKFLMTDERTKRIPKYLETPGGEN
WQKEIGELLKFSKNRDS"
complement(4202..4378)
/gene="CP0015"
CDS
complement(4202..4378)
/note="CP0015"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF37911.1"
/db_xref="GI:7188954"
/translation="MSFLLIKFSPPLKQDYLFDISPYTSEITIGSYFKLNKASL
QSSTLRKRSIIS"
4362..6005
/gene="CP0016"
CDS
4362..6005
/note="CP0016"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37912.1"
/db_xref="GI:7188955"
/translation="MSRKDNVSLARSIFNLSGTFCSRITGIFREIATAMATYFADPI
VAAWFLGRTVFELKILGLQAFIPHEFLRAQSLDRAAFRRFSLIKGSTI
LFTLLIEAVLVVLYVEGYDMLLTMLPCGIFLMYVNGALLHCENKFFGVG
LAPVYVNIWIFFVIARHSDPRERIGLSVALVIGFFFEWLITVGVWKLLEAKSP
FOEHSVRALLAPLSGLTSSIFOLNLSIDICLARYVHEIGPLYLMYSLKIYOLPIH
LFEGFVFTVLPAISRCVQREDHERGLKMFVLTMTSMVMIIMTAGLLILLLEGVRY
LYEHGLFPQSAVAIVRYLGASIIIPMALAPLVSVLFYARQYAVPLFTIGITALA
NIVLSLGRWLKDVSGISVATSIITAVOLYELWYSSKRLPMYKSLWESIRRSIK
VWGTMLACMTLGNLITQTYVIFLNPPLPLANPLSSITQAIAFLSCEFLAFI
FGPAKLLRVDELINLASFEYWRGQGLLQROHWMQDTQN"
6108..7376

```

```

/gene="CP0017"
6108..7376
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37913.1"
/db_xref="GI:7188956"
/translation="MKQVYQWLASVLLATISGYAELPLSEQKVKSHVTITLDEVK
DYLKRGFVETRQKQDGLRIAGDVRARWLYERDINKPSDFKDYKPNLPVNRSEFYI
YIDYABRNWLSKMNMTAAGENTAAGVINDIRAFGLRGYRKNPRTDFEFMIGRS
GLGLDFEVOFQSNDFGLHIYWTRELSKDYPOYVIVHGGPEVVMYTKKHVAWVEGI
LNRLPKQFFVKCSVDWMTFVPSSTTEKATNAMYKYCVQWOLYKHKHQSQVPIG
OKPLYLYGAFLMPLAKATKTTINGENLAWFTGGTLGLGRKAGDSATVRYEYEA
LSVPEIDVSGIGRGNLLKFWFAQIAANYDPKEANGFTNYKGFSAIYMYGITSLSFR
7488..9443
/gene="CP0018"
CDS
7488..9443
/note="CP0018"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37914.1"
/db_xref="GI:7188957"
/translation="MVNPIGPIDEITERTPPADLSAOGLEASANKSAEORAGAE
AKPKESKTSVERMSILRSAYNALMSLADKLGIASSNSSTSRSDVDTATATPT
PPPTFDYQKTOAYDTIFTSTLADIQAALVSLQDAVTIKDTAADETAIAAEW
ETKANADVCAQITELAKYASDNOAILDLSGKLTSPDLOAALLOSVANNNKAELL
KEMODNPVPKTFAIAGSLVDQDATATQTEKGNALRDAYFAGORASGAVENAKSN
NSINISAKAIAATATQIAEAKKFPDPSILQEAQWVIOAKDLNKAIRPADGSDV
PNPTGTGSKQOQSGISIRVSMLLDADENETASILMSGFRQIMHMTNPNDSQA
OOELAAQAAKAGDDSAALADAQALKAALGKAGQOQGINALGQRIASAAVVA
GVPPAAASSIGSSVKQLYKTSKTSQDYKTOISAGYDAYKSIDNAYGRANDATRDVI
NNVSTPALTRSVPRARTFARGPKTDQALARVISGNSRTIGDVTSOVSALOSVMOIIO
SNPOANNEIRQKLTSAVTKPPQGYPVQLSNDSTQKFIAKLESFAEGSRTAAEIK
ALSFTNSLFIQQLVNLNIGLSYGLQ"
complement(9770..12388)
/gene="CP0019"
CDS
complement(9770..12388)
/note="CP0019"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37915.1"
/db_xref="GI:7188958"
/translation="MSFESIONRLTISGESTRIKLDHKYSGDPRSPVPAINLEELN
SGIYALRHLNMALOSENFVAALLNPNNTIPTTSWTDYKHSRQOASSPRAPSTPT
DIVSAAALALVLDVGGELVAVSVEIDLGALSTIVRQLMASYLGLTTLTAQEK
VVFSSYVPSEKLLHVKQKAEIKQAEIKAVLEAKGVSTEEIAILKEYPDYI
AADFEKEFIEPLRYAKVCAPIQEMNENAIQLPTTPTPDNNVNGMNTLSKI
LOAIDDAIKQAPALGGQDEIITILQTLVPLVDKTFKAEFDLLYATQLNTASIKL
YLTDQTAIEYRGKITKYQNSIQNLSETRKRVNRRSMLETQLSMFQOQNCFTWIS
QANALNATITNKYISAVLTSMEMVGGLLCLSYMERLADDEKAIKFKSVNEYLPIHI
VQGSWNGWIAKMAAYQELAEYSLGTAVTSDQIKAYLOTGRNEFKATRHFFHIGD
OMYOFANETVFNCLITWANGAIQPDLGFIREAMTVNGTVEADYVSNAAORTLNEFNTA
ATAVHLOLOIAELOKADDDLPKGAFTENRKFVAVAWITTSLSGALLISMLNSQ
LPKQEAFLKPLIEENFNLANALNSLLOITNEFTTSVYSSLSYLVSQSKTGQNLF

```

Query Match

92.5%; Score 1359.4; DB 1; Length 12548;

: Best Local Similarity 99.5%; Pred. No. 0;

Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 1 atgggtaatactcattgggtccaggtcctatagacgaacacacacacctcccgagat 60

|||||

Db 7488 ATGGTTAATCCTATTGGTCCAGGTCCTATAGACGAACACACACCTCCCGAGAT 7547



QY 61 cttctgtcaaggattgagcagtgagcagcaataaagagtgaggaaagctcaagaata 120  
DB 7548 CTTTCTGCTCAGGATTTGGAGCGAGTGCAGCAATAGAGTGCAGAGCTCAAGATA 7607  
QY 121 gcaggtgcggaagctaaagcctaaagaattcaagaccgattctgtagagcgatgagc 180  
DB 7608 GCAGGTGCGGAAGCTAAGCCTAAAGAACTTAAGACCGATTCTGTAGAGCGATGAGCATC 7667  
QY 181 ttgcgttgcagtgaaatgctctcagtgagctggcagataagctgggtattcttctagt 240  
DB 7668 TTGCGTTTTCGAGTGAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT 7727  
QY 241 aacagctgtctctactagcagatctgcagacgtggactcaacacagcagcagcact 300  
DB 7728 AACAGCTGCTTCTACTAGCAGATCTGCAGACGTGGACTCAACACAGACGACCGCAGCT 7787  
QY 301 acgctctctccaccacgtctgatattataagactcaagcgcaaacagcttaagatact 360  
DB 7788 ACGCCTCTCCACCACGCTTTCATGATTATAAGACTCAAGCGCAAAACAGCTTACGATACT 7847  
QY 361 attcttaacctcaacatcactagctacatacagagctgttggtagcctccagatgct 420  
DB 7848 ATCTTTACCTCAACATCACTAGCTGACATACAGGCTGCTTTGGTAGGCTCCAGGATGCT 7907  
QY 421 gtcactaatataaagatacagcgctactgatgaggaacccgcaatcgctgcgagtg 480  
DB 7908 GTCACTAATAAAGATACACGCGCTACTGATGAGGAACCGCAATCGCTGCGAGTGG 7967  
QY 481 gaactaagaatgcgcagtgcaattaaagtggcgcgcaaatatacagaattagcgaaatat 540  
DB 7968 GAAACTAAGAATCGCGATGCAGTTAAAGTTGGCGCGCAAAATACAGAAATAGCGAAATAT 8027  
QY 541 gcttcgataaaccagcaattctgaactctttagataaactgaactctctgcacctcta 600  
DB 8028 GCTTCGATACCAACGCGATCTTGACCTTTAGGTAACTGACTTCCTTCGACCTCTTA 8087  
QY 601 cagactgctctctccaatctgtagcaacaataaacaagcagctgagctctttaaagag 660  
DB 8088 CAGGCTGCTTCTTCCAACTCTAGCAACAATAAACAAGCAGCTGAGCTTCTTAAAGAG 8147  
QY 661 atgcaagataaaccagtagtccaggggaaacgcctgcaattgctcaattcttagttgat 720  
DB 8148 ATGCAAGATAAACCCAGTAGTCCAGGAAACGCTGCAATTCGTCAAATCTTTAGTTGAT 8207  
QY 721 cagacagatgctacagcgacacagatagagaagaatggaatgcgattgggagtcacat 780  
DB 8208 CAGACAGATGCTACACGCGACACAGATAGAGAAGATGGAATGCCATAGGGATGCATAT 8267  
QY 781 ttgcagcagagaacgctagtgtgagctgtagaaaatgctaaatcttaataacagtaagc 840  
DB 8268 TTTCAGGACAGAACGCTAGTGGAGCTGTAGAAAATGCTAAATCTTAATAACAGTATAAGC 8327  
QY 841 acatagatcagctaaagcagcaatcgtactgctgaagacaacaatagctgaagctcag 900  
DB 8328 ACATAGATTACGTTAAAGCAGCAATCGCTACTGCTAGACACAATAGCTGAAGCTCAG 8387  
QY 901 aaaaagtcccgactctccaatcttccaagaagcggaacaatgtgtaatacaggtcag 960  
DB 8388 AAAAAAGTTCCCGACTCTCCAACTTCTTCAAGAAGCGGAACAATGTAATACAGGCTGAG 8447  
QY 961 aaagattcttaaaaataatacaactcagatggtctctgattgtcccaataccaggaactaca 1020  
DB 8448 AAAGATCTTAAAAATATCAAACTGCAGATGTTCTGATGTTTCCAAATCCAGGAACACTACA 8507  
QY 1021 gttgaggtcccaagcaacaagaagtagtattgttagtattcggtttccatctgttta 1080  
DB 8508 GTTGGAGGCTCCAGCAACAAGAGTAGTATTGTTAGTATTTCGGTTCCTTCATGCTGTTA 8567  
QY 1081 gatgagctgaaataagagaccgtctccattttgatctgtggttttcgtagatgattcac 1140  
DB 8568 GATGATGCTGAAATAGAGACCGCTCTCCATTTTGTATGCTCTGGTTCCTGTCAGATGATTCAC 8627  
QY 1141 atgttcaatacggaaataatctgattctcaagctgccccacaagagctcgcagcaagaact 1200

DB 8628 ATGTTCAATACGGAATAATCCTGATTCTCAAGCTGCCCAACAGGAGCTCGCAGCAAGCT 8687  
QY 1201 agacacggaagccgctggagatgacagtgctgctcagcgctggcagatgctccagaaa 1260  
DB 8688 AGACGACGAAGCCGCTGGAGATGACAGTGTCTGCTCAGCGCTGGCAGATGCTCAGAAA 8747  
QY 1261 gcttaagcggcctctaggtaaagctgggcaacaacagggcactactcaatgctttggga 1320  
DB 8748 GCTTTAGAGCGGCTCTAGGTAAAGCTGGGCAACAACAGGGCATACTCAATGCTTTAGGA 8807  
QY 1321 cagatcgcttctgctgctgtgtgagcgaggag-tctcccgctgcagcaagtcttatg 1379  
DB 8808 CAGATCGTCTCTGCTGCTGTTGTTGAGCGCAGGAGTTCTCCCGCTGCAGCAAGTCTTATA 8867  
QY 1380 g 1380  
DB 8868 G 8868  
  
RESULT 6  
AP002547/c  
LOCUS AP002547 300550 bp DNA BCT 08-DEC-2000  
DEFINITION Chlamydomphila pneumoniae J138 genomic DNA, complete sequence, section 3/4.  
ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008  
VERSION AP002547.2 GI:10176693  
KEYWORDS  
SOURCE Chlamydomphila pneumoniae J138 (strain:J138) DNA.  
ORGANISM Chlamydomphila pneumoniae J138  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomphila.  
REFERENCE 1 (sites)  
AUTHORS Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.  
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA  
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
MEDLINE 2030349  
REFERENCE 2 (bases 1 to 300550)  
AUTHORS Shirai, M.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases. Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail: mshirai@po.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227, Fax: 81-836-22-2415)  
COMMENT On Sep 15, 2000 this sequence version replaced gi:6172298 gi:6172300 gi:6172396 gi:6172398 gi:8978889 AB033786-AB033787, AB033816-AB033817: Submitted (25-Oct-1999).  
FEATURES  
source  
1. 300550  
/organism="Chlamydomphila pneumoniae J138"  
/strain="J138"  
/db\_xref="taxon:138677"  
/note="synonym:Chlamydia pneumoniae (strain J138)"  
complement(21..794)  
/gene="dapF"  
complement(21..794)  
/gene="dapF"  
/codon\_start=1  
/transl\_table=11  
/product="diaminopimelate epimerase"  
/protein\_id="BAA98725.1"  
/db\_xref="GI:8978890"  
/translation="MAFVSPSTISKYFTYSGAGNRLLETLPVEDVRFLOCTRVDFGLYLPSSCADAOLIIFNDSGSRPTGCGNLCRAIAHLSOKCKSDISVSTDSGLYSGYFSWDRVLVDWTLADWRASVHRLSRPDLPKREVCIHTGVHAYVILPEISTLDLSILGFLRYHOTFPDGVNPNVQILGLCHQLKRYTERGVEGTAAACGTGALASLVVSNYSGWKESTQIHTWGGELMTVSQNRGRVYLQGSVTRDL"  
complement(763..1338)  
/gene="clpP\_1"  
complement(763..1338)  
/gene="clpP\_1"



```

/codon_start=1
/transl_table=11
/product="CLP protease"
/protein_id="BAA98726.1"
/db_xref="GI:8978891"
/translation="MADGEVHKLRDIEKELLEARRVEFFSEPTVEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDIKMLTSPVTVTYVGLAASGWSVLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKADRMWMTANAKDFGLDGLIFSEFNDL"
complement(1358. .2917)
/gene="glaA"
/complement(1358. .2917)
/gene="glaA"
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/translation="MLKVFEKPKPAIVEIFTKVAVVSVLLHKLLENASGKKGOSLAS
TAYLAALDHLNAPSGIGERIIDEKLSQRSHLKMIASENYSSLSVQLAMGNLLTDKYC
EGSPFKRYSCYENVDATIEWECVETAKELFAADCACQVPHSGADHLAVNAIILTHKV
QGPVSKRYKVTNELTEETVTLKAEMSSCVCLGPSLNSGGHLLTHGNVRLNVMSKLM
RCFPVNDPDECTDYAEISRLAKKEYPKVLIAGYSSYSRRLNFAVLKQIAEDCGSVL
WYMAHRAGLVAGGVFDEBNPIYADIVTTTTHTKLRGPRGLVLAETREYESTLUNKA
CPLMGGPLPHVIAAKTVALKEALUSDPKKYAHQVNNARLAEFLSHGURLLTGGT
DNHMVIDLGSIGTSGKTAEDILSGVGVNRRNSLPSDAIGKWDTSIGRLGTPALTTL
GMGIDEMEVEDADIIVKVLNRILRSCHVEGSSKKNKGELPEAIAQEARDRVRLNLLRFP
LYPEIDLEALV"
3107. .3778
/gene="CPj0522"
3107. .3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/translation="MTLYLGLNQTKARKYQAHYLPILFLFYAKSTPONKRALQFLPQ
ATHVLTPSTSHFLSRMTSLLSKATLTKTYLIGESTKERLLSFLGQVKYVYVATQ
ETAEIGFLLQALPSSARILYPHSLARPVIREFLYNRFTFSYPHYTVKPKLKNKI
LSYKKIIFTSPSTVRAFAPKIFPRPEPTYWCQGRMTLQEFQKFSQKQVSLLETIGK
SRTSP"
3843. .4175
/gene="CPj0523"
3843. .4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/translation="MASSATPGDGTAPSLFPATPRPNRYNFKLALFVTTAIALVWIAL
IATTTAIGLCIHLCSFELTAIPLYFISRYICSHYARNVYIALDVVDFHSLQDMRS
HSPIFSDF"
4223. .5302
/gene="CPj0524"
4223. .5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/translation="MSPGSRTESSQVSVLSPVRDKETAPKQFTIAKISTIALIASL
ALGALVAGISITIVLGNPVFLALITTFVSVTVFLVYHOMTSKYSSNWKVLEONFK
PLGRAWQEKNDVCSNEMQVFNHLPKFAIOTDASQFPQFTLGLRVEKNQST
GILFNPVGTNLONTATNLSTILYSLTKLSYMDTCKQREGGAKGDEPSPTEYRV
KALPNEALDQFNLLNSAERKSILPFLGHVCGPKSEELPNQOYIYQOALLAYENCL
VAPRESHAALVALPLFTSVYEPPEELPKEGTFYWDNQTOAFCKRALLDAIQNTALR
YPORESVILQDPNTIESQSRSE"
5642. .6406
/gene="CPj0525"
5642. .6406

```

```

/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="CT398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/translation="MHDALLSLAIQELDIKIMRLMRVYKKEHQKELAKVQSLKSDIR
KVQKLEMLNKTQIRDGENRIQISEQINKLENOQAQVKKMDFNALTQEMTTANK
ERRSLHOLSMDKQAGEDLIVLSKESLASTENSSSVIEKEIFESIKKNEEGKAL
LEORTELKATNPILLSIYERLLANKKDRVVVPTENRVCSCHIVLTPOHENLYRKKD
RLIFCEHCSRILYQWESVYNAQENSTARRRRAAV"
complement(6827. .7816)
/gene="kpsF"
/complement(6827. .7816)
/gene="kpsF"
/codon_start=1
/transl_table=11
/product="GutQ/kpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/translation="MPSPMISTDVCODILGKQKEAVDFFFOAFQPKAMOLAELTLGH
SGWFFSGVSGKGVARKLVATLQSLSEFALFFSPVDLLHGDGLGVSPGDIIVCLFSKS
GETOELLDTFHLKSRRAILVAITSMPTSNLAUSDVILVPSVAELDPPFNLIPTNST
TCQMFIDGLFAMLLFPHSGVSLSTYKNGHPSGQVGMKANGKVKQDFMFKTEVPFCHLG
DKV3FSLEFVAYCGCVCIYDPOFLMGIFTDGLRRSLASVYGDGCVLSLSLEKYNTA
NPRCTEDSDIAIALQLMESSPVAVLPVLDNEENRHVTGLLHMHTLAKAGLL"
complement(7843. .9024)
/gene="sucB_2"
/complement(7843. .9024)
/gene="sucB_2"
/codon_start=1
/transl_table=11
/product="dihydrolipoamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/translation="MIFEFPPKIGETSSGGSIVRWLNGLGDHVARDPLEIVSTDKI
ATEPLPKAGRLVRFCVNEGDEVDGLVLEIEISEADEDSCPLTSCETKSEA
GSSSSWFSFPAVLSLAQREGIDLNQKIACTGKGRVTRQDLLEYISEQQVSIPE
IFQEVNRIKMSPLRRALIASLSKSSDEVPASIVVDVTDLMNLISGERQRLDTH
NKDITFSFIVQCLAOTRQFPLNGSLDGTIVMKKSNVNGVAVNLNKEGVVVPVH
NCQDGLVLSAKALADLSRARLNKLDSPEDQDSVTVTFNGMTGALIGMPIIRYPEV
AIIIGTITQKRVVRRDDSLAIRKVVYVTLTFDHRVLDGIYGEFSLTKNLSVVM
G"
complement(9038. .10282)
/gene="gltA"
/complement(9038. .10282)
/gene="gltA"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/translation="MKLMMKIFIGLVGVTLGLVLEDKAIFPKPIGIDIFLNLLSVVY
PLVCSWVIGTASIDMKKRLGKSVGLYGLTITATAIIVIGLCFAWIFSPNGCDFA
QAQMSDASVTVIDSNKTAAYFLSIIAQVFPSPNVSFAEGNLIQIIIFAIFLGIALRL
SGERGRPVREIDGSETMLRMVNMIMSFPYGVGASMAWISGNHGLGMLQGLKFI
AYLACLAFHATLVFGLGVFCGKMSFSLSSMDDAISCAVSSSSATLPVTRCVS
KNLGVSAEVSGFVLPLGATVNMNGTAIFQGMAAVTAQAYNCPLSLSLLLVVTATF
SAVGSAGVPGGGMITLGSVLASVGLPIQIALLAGIDRLRDIIVCTPMILGDVAVVATY
VASEGELSPYESIKQESVET"
complement(10282. .11379)
/gene="ycaH"
/complement(10282. .11379)
/gene="ycaH"
/codon_start=1
/transl_table=11
/product="ATPase"
/protein_id="BAA98735.1"
/db_xref="GI:8978900"
/translation="MKRRFPSTFLFYRRVTTAISLEGILGWGLGSLLSKVAFLVA
CWNRFWSVTPYRSTVTSVGNIVVGAGKPTVLMALRALRGLCYSCGLSRGVKSO
SSROKLTVDKSKVHSASVYGDPELLMAEKLPYEGSVWYHKDRRISAAAEKFGILL
DDGLQYRKLHKDVEIAVNVGQDPIGGRAFFPKGRULRDFPLRLKTVDAIIVNGGKREAG

```



TVVRVSNAPQIFVKPTIASVWTHNGERIPKEALRELRVGVFCGLGPGQFGLNMLRE  
EGIHGLKYLPLDRAHATKKELNYFCOAMROGGLCTEKDSVKLPRLSGEVSLLP  
IAKVEMLSVNQDRTLILLNNIEQHKNRGN"

Query Match	92.5%	Score 1359.4	DB 1	Length 300550
Best Local Similarity	99.5%	Pred. No. 0		
Matches 1374	Conservative 0	Mismatches 6	Indels 1	Gaps 1
Qy 1	atgggttaattcctattggtccaggctctatagacgaacacagacacacacccctccgcagat	60		
Db 217598	ATGGTTAATCCCTATTGGTCCAGGCTCTATAGACGAACAGACGACACCTCCCGCAGAT	217539		
Qy 61	ctttctgctcaaggattgagcgcagtgacgaataaagtgagcgaagctcaagaata	120		
Db 217538	CTTTCTGCTCAAGGATTGGAGCGAGTGCAGCAATAAAGATGCGGAAGCTCAAGAATA	217479		
Qy 121	cgaggtgcgaagcttaagcctaaagaattcaagacgcgattctgtagacgctgagacatc	180		
Db 217478	GCAGGTGGGAAGCTTAAGCCTTAAGAAATCTAAGACCGATTCTGTAGACGATGGAGCATC	217419		
Qy 181	ttgcgtctgcagtgatgctctcattgctgagctggcagataagctgggtattgctctagt	240		
Db 217418	TTGCGTCTGCAAGTGAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT	217359		
Qy 241	aacagctogtcttactagcagatctgcagacgtggactcaacacagcagcagcagct	300		
Db 217358	AACAGCTGCTCTTCTACTAGCAGATCTGCAGACGTGGACTCAACACAGCAGCGCACCT	217299		
Qy 301	agcgtctctccaccagctctgattataaagaactcaagcgaagcgaacagcttacgatact	360		
Db 217298	AGCGCTCTCTCCACCCACGCTTTGATGATTATAAGACTCAAGCGCAACACAGCTTACGATACT	217239		
Qy 361	atcttaacctcaactcagctagctgacatacagctgctcttggtgagcctcagatgct	420		
Db 217238	ATCTTTACCTCAACATCACTACTACTGACATACAGGCTGCTTTGGTACGCTCCAGGATGCT	217179		
Qy 421	gtcactaatataaaggatcacagcgcctactgatgaggaaacgcgaactcgcgtcgagtg	480		
Db 217178	GTCACTAATATAAAGGATACAGCGGCTACTGATGAGGAACCGCAATCGCTCGGAGTGG	217119		
Qy 481	gaactaagaatgccagtgcaattaaagttggcgcgaataattacagaattagcgaatat	540		
Db 217118	GAACTAAGAATGGCGATGCGATTAAGTGGCGCGCAATATACAGAATTAGCGAAATAT	217059		
Qy 541	gcttcggataaacaagcagcttctgactctttgagtaactgaactcctctcgaactctta	600		
Db 217058	GCTTCGGATAACCAAGCGATTCTTGACCTTTAGTAACTGACTTCCTTCGACCTCTTA	216999		
Qy 601	cagactgctctctccaatctctagcaacaataaacaagcagctgagcttcttaagag	660		
Db 216998	CAGGCTGCTCTTCTCCAACTCTGTAGCAACAATAACAAGCAGCTGAGCTTCTTAAGAG	216939		
Qy 661	atgcaagataaccagtagtccagggaaaacgcctgcaattgctcaattcttagttgat	720		
Db 216938	ATGCAGATAAACCCAGTAGTCCCGGGAACCGCTGCAATTGCTCAATCTTTAGTTGAT	216879		
Qy 721	cagacagatgctacagcagacagatagagaagatggaatgcgattgggagatcatat	780		
Db 216878	CAGACAGATGCTACAGCGACACAGATAGGAAGATGGAATGCCATAGGATGCATAT	216819		
Qy 781	tttcgaggacagaaacgctagtgagctgtagaaaaatgcataatcaataacagtagtaagc	840		
Db 216818	TTTGCAGGACAGACGCTAGTGGAGCTGTAGAAAATGCTAAATCTAATAACAGTATAAGC	216759		
Qy 841	acatatagctcagctaaagcagcaatcgtactgctgaagacacaataagctgaagctcag	900		
Db 216758	AACATAGATTACGTAAAGCAGCAATCGCTACTGCTAGACACAATAAGCTCAG	216699		
Qy 901	aaaaagttcccgactctccattctcaagagcgggaacaaatggttaatacagctcag	960		
Db 216698	AAAAGTTCCCGACTCTCCAACTCTTCAAGAGCGGACAAATGTTAATACAGGCTGAG	216639		
Qy 961	aagatctttaaaaaatatcaaacctgcagatggtctgtatgttctccaaatccaggaactaca	1020		

Db 216638	AAAGATCTTAAAAATATCAAACTGCAGATGTTCTGATGTTCCAAATCCAGGAACACTACA	216579		
Qy 1021	gttcgagctccaaagcaacgaagtagtattgtagtattcgtgttccatcgtctgta	1080		
Db 216578	GTTGAGGCTCCAAGCAACAAGGAAGTAGTATTGTTAGTATTGTTGTTTCCATGCTGTTA	216519		
Qy 1081	gatgatgctgaataatgagacgcctccatttgcgtctgggttgcgtcagatgattcac	1140		
Db 216518	GATGATGCTGAATATGACCGCTTCCATTGTTGATGCTGGTTTCGTCAGATGATCAC	216459		
Qy 1141	atgttcaatacggaaaaatccttgattctcaagctgcccccaacgagcgtgcgcagcaagct	1200		
Db 216458	ATGTTCAATACGAAAAATCTGATTCTCAAGCTGCCAACAGGAGCTCGCAGCACAGCT	216399		
Qy 1201	agacagcgaagccgctggagatgacagtgctgctgcagcgtggcagatgctcagaaa	1260		
Db 216398	AGACAGCGAAGCGCTGGAGATGACAGTGTCTGTCAGCGCTGGCAGATGCTCAGAAA	216339		
Qy 1261	gcttttagaagcggctctaggtaaagctgggcaacacagggcgcatactcaatgctttggga	1320		
Db 216338	GCTTTAGAAGCGCTCTAGGTAAAGCTGGGCAACAACAGGGCATACTCAATGCTTTAGGA	216279		
Qy 1321	cagatcgcttctgctgctgtgtgagcgcaggag-tcctcccgctgcagcaggtctctatg	1379		
Db 216278	CAGATCGCTTCTGCTGCTGTTGTGAGCGCAGGAGTTCTCTCGCTGCAGCAAGTTCTATA	216219		
Qy 1380	g 1380			
Db 216218	G 216218			
RESULT 7				
AX045135				
LOCUS	AX045135	1456 bp	DNA	PAT 24-NOV-2000
DEFINITION	Sequence 5 from Patent WO0066739.			
ACCESSION	AX045135			
VERSION	AX045135.1	GI:11343734		
KEYWORDS				
SOURCE	Chlamydia pneumoniae.			
ORGANISM	Chlamydia pneumoniae			
REFERENCE	1 (bases 1 to 1456)			
AUTHORS	Murdin,A.D., Omen,R.P., Wang,J. and Dunn,P.			
TITLE	i(chlamydia) antigens and corresponding dna fragments and uses thereof			
JOURNAL	Patent: WO 0066739-A 5 09-NOV-2000;			
FEATURES	Aventis Pasteur Limited (CA)			
source	Location/Qualifiers			
	1..1456			
	/organism="Chlamydia pneumoniae"			
	/db_xref="taxon:83558"			
CDS	101..>1452			
	/note="unnamed protein product"			
	/codon_start=1			
	/translation="MVPNIGPGPIDETERPPADLSAQGLEASANKSAEQAQTAGAE			
	AKPKESITDSVERWSILRSVAVNALMSLADKLGIASSNSSSSSTSRSDVDSTATPT			
	PPPTFDYKTAQATYDTIFTSTLADTQALVSLQDAVTNFKDTAATDEETAIAAEW			
	ETKNADVKKVGAQITELAKVASDNQAILDLSLGTCTFDLLQALQLSVANNKAAELL			
	KEMODNPVPGKTPAIQSLVDOTDATOLEKDGNAITRDYAFAGASGAVENAKSN			
	NSISNIISAKAATATATKTQIAEAKKFPDPIQEAQEQMTOAEKDLKNIKPAGSDV			
	PNQTTVTGGSKQGGSSIGSRVSMLLDDAENETSLMSGFRQMHMFNTENPDSQAA			
	QOELAAQARAARAAAGDDSAALADQAALGAALGKAGQGGGILNALGQIASAAVWSA			
	G"			
BASE COUNT	452 a 308 c 331 g 365 t			
ORIGIN				
Query Match	91.6%	Score 1347	DB 6	Length 1456
Best Local Similarity	99.6%	Pred. No. 0		



Matches	1350;	Conservative	0;	Mismatches	5;	Indels	0;	Gaps	0;
---------	-------	--------------	----	------------	----	--------	----	------	----

QY	1	atggttaatcctattggtccagggtccctatagacgaaacagaaacgacacacctccgcagat	60
DB			
QY	61	ctttctgctcaaggattggaaggcgagtgccagcaataaagagtgcggaagctcaaaagata	120
DB			
QY	161	ctttctgctcaaggattggaaggcgagtgccagcaataaagagtgcggaagctcaaaagata	220
DB			
QY	121	gcaggtgcggaagctaaagcctaaagaaacctaaagaccgattctgtagagcgatggagcacc	180
DB			
QY	221	gcaggtgcggaagctaaagcctaaagaaacctaaagaccgattctgtagagcgatggagcacc	280
DB			
QY	181	ttgcgtcttgcaagtgaaatgctctcatagctgcgcagataaagctggggatattgcttctagt	240
DB			
QY	281	ttgcgtcttgcaagtgaaatgctctcatagctgcgcagataaagctggggatattgcttctagt	340
DB			
QY	241	aacagctcgctctctactagcagatctgcagacgtggactcaacgacgacgacgacgacct	300
DB			
QY	341	anacgctcgctctctactagcagatctgcagacgtggactcaacgacgacgacgacgacct	400
DB			
QY	301	agcgtctctccaccacgctctgattattataagactcaagcgcaaaacagettacgatact	360
DB			
QY	401	acgctctctccaccacgctctgattattataagactcaagcgcaaaacagettacgatact	460
DB			
QY	361	atctttactcaaacatactactagtcgaatacagactccttggtagacctccaggaatgct	420
DB			
QY	461	atctttactcaaacatactactagtcgaatacagactccttggtagacctccaggaatgct	520
DB			
QY	421	gtcactaataaaggatcacagcggtactgtaggaaacgcaactcgctgcggagtg	480
DB			
QY	521	gtcactaataaaggatcacagcggtactgtaggaaacgcaactcgctgcggagtg	580
DB			
QY	481	gaaactaagaatgcgagtgcaattaaagttggcgcgcaaatattacagaatttagcgaatat	540
DB			
QY	581	gaaactaagaatgcgagtgcaattaaagttggcgcgcaaatattacagaatttagcgaatat	640
DB			
QY	541	gcttcggataacccaagcgattcttgactccttagttaaacgactccttcgcaccttta	600
DB			
QY	641	gcttcggataacccaagcgattcttgactccttagttaaacgactccttcgcaccttta	700
DB			
QY	601	cagactgctctctccaatctgtagcaaaacaataacaagaacgctgagctcttaagaag	660
DB			
QY	701	caggctgctctctccaatctgtagcaaaacaataacaagaacgctgagctcttaagaag	760
DB			
QY	661	atgcagaataaccccgtagtcccagggaaacgcctgcaattgctcaactcttagttgat	720
DB			
QY	761	atgcagaataaccccgtagtcccagggaaacgcctgcaattgctcaactcttagttgat	820
DB			
QY	721	cagacagatgctacagcgacacagatagagaagaatggaaatgcgattgggggatgcata	780
DB			
QY	821	cagacagatgctacagcgacacagatagagaagaatggaaatgcgattgggggatgcata	880
DB			
QY	781	tttcgaggacagaaacgctlagtggagctgtagaaaaatgctaaactaataacagataaagc	840
DB			
QY	881	tttcgaggacagaaacgctlagtggagctgtagaaaaatgctaaactaataacagataaagc	940
DB			
QY	841	aacatagattcagctaaagcgacactcgctactgtaagacacaaatagctgaagctcag	900
DB			
QY	941	aacatagattcagctaaagcgacactcgctactgtaagacacaaatagctgaagctcag	1000
DB			
QY	901	aaaaagttccccgactctccaatcttcaagaagcgcaaaaatggttaaatacaggctgag	960
DB			
QY	1001	aaaaagttccccgactctccaatcttcaagaagcgcaaaaatggttaaatacaggctgag	1060
DB			
QY	961	aaagatctttaaaaaatcaaacacctgcagatggttctgattgtcccaaatcaggaaactaca	1020
DB			
QY	1061	aaagatctttaaaaaatcaaacacctgcagatggttctgattgtcccaaatcaggaaactaca	1120
DB			
QY	1021	gttgaggctccaagcaacaacgggaagtagtatttggttagtattcggttccatgctgatta	1080
DB			
QY	1121	gttgaggctccaagcaacaacgggaagtagtatttggttagtattcggttccatgctgatta	1180
DB			

[illegible]



Db 61 TCTCAGACGTTGACTCAACACACAGCAGCAGCAGCTAGCGCTCTCTCCACACCGCTTTGAT 120  
Qy 325 gattataagactcaagcgcaaaagcttaacgataactatcttttactctcaacatcaactagct 384  
Db 121 GATTATAAGACTCAAGCGCAACAGCTTTACGATACACTATCTTTTACTCTCAACATCACTAGCT 180  
Qy 385 gacatacagggctgctttgttgtagctccagagatgctgctcactaataataaagatcacagcg 444  
Db 181 GACATACAGGCTGCTTTGGTGAGCCTCCAGGATGCTGTCACATAATATAAGAGTACAGCG 240  
Qy 445 gctactgatgagaaacccgaatcgctgagtgaggagtggaacactaagaatgcogagtcgaatt 504  
Db 241 GCTACTGATGAGAAACCGCAATCGCTCGGAGTGGGAAACTAAGAAATGCCGATCGAGTT 300  
Qy 505 aaagttgagcgcaaaattacagaattagcgaaatattgcttgcgataacccaagcgtatctt 564  
Db 301 AAAGTTGCGCGCAAAATACAGAAATAGCGAAATATGCTTCGGATAACCAAGCGATTCCT 360  
Qy 565 gactcttttagttaactgactctcttcgacctcttaeagactgctcttctccaatctgta 624  
Db 361 GACTCTTTAGTAAACTGACTCTCTTCGACCTCTTACAGGCTGCTCTTCTCCAATCTGTA 420  
Qy 625 gcaacaataacaaagcagctgagcttcttaagagatgcaagataaccccgtagtccca 684  
Db 421 GCAAAACAATAACAAAGCAGCTGAGCTTCTTAAAGAGATGCAAGATAACCCAGTAGTCCCA 480  
Qy 685 gggaaaaacgctgcaattgctcaatcttttagttgatcagacagatgctacagcgacacag 744  
Db 481 GGGAAACGCTTGAATTTGCTCAATCTTTAGTTGATCAGACAGATGCTTACAGCGACACAG 540  
Qy 745 atagagaaagatggaatgcgattgggagtcataatttttgaggacagaaacgctagtgga 804  
Db 541 ATAGAGAAGATGGAATTCGATTTAGGATGCAATATTTTGAGGACAGACGCTAGTGA 600  
Qy 805 gctgtagaaatgctaaatcttaataacagatataagcaacatagattcagctaaagcagca 864  
Db 601 GCTGTAGAAATGCTAAATCTAATAACAGTATAAAGCAACATAGATTACGCTAAAGCAGCA 660  
Qy 865 atcgctactgtaagacaaatagctgaagctcagaaagttcccgactctccaatt 924  
Db 661 ATCGCTACTGCTTAAGACAAATAGCTGAAGCTCAGAAAGATTTCCCGACTCTCCAATT 720  
Qy 925 cttaagaagcggaacaaatgtaatacagcgctgagaagatctttaaaaaatatacaacct 984  
Db 721 CTTCAAGRAGCGGAACAATGTAATACAGGCTGAGAAAGATCTTAAAAATATCAAACT 780  
Qy 985 gcagatggtctgattgttccaaatccaggaactacagttgaggtcccaagcaacaagga 1044  
Db 781 GCAGATGTTCTGATGTTCCAAATCCAGGAACTACAGTTGGAGGCTCCAAGCAACAAGGA 840  
Qy 1045 agtagtattgtagtattgctttccatgctgttagatgactgaaatgagaccgct 1104  
Db 841 AGTAGTATTGGTAGTATTGCTTTCCATGCTTTAGATGATGCTGAAATGAGACCGCT 900  
Qy 1105 tccattttgattgctgggtttgctcagatgattccatgttcaatcacggaataatcctgat 1164  
Db 901 TCCATTTGATGCTGGGTTTCGTGATGATTCATGTTCAATACGGAATTCCTGAT 960  
Qy 1165 tctcaagctgccaacagagctgcagcacaagctagacgagcgaagccgctggagat 1224  
Db 961 TCTCAAGCTGCCCAACAGGAGCTCGCAGCACCAAGCTAGACGAGCGAAAGCCGCTGGAGAT 1020  
Qy 1225 gacagtgcctgcagcgctgcaatgctcagaagctttagaagcgcttaggttaa 1284  
Db 1021 GACAGTGTCTGCAGCGCTGCAGATGCTCAGAAAGCTTTTAAAGCGGCTCTAGGTAAA 1080  
Qy 1285 gctgggcaacaacagggcactaatgctttgggacagatcgcttctgctgctgttgg 1344  
Db 1081 GCTGGGCAACAACAGGGCATACTCAATGCTTTTAGGACAGATCGCTTCTGCTGCTGTGTG 1140  
Qy 1345 agcgcaggag-tcttccccctgcagcaagttctatgg 1380  
Db 1141 AGCGCAGGAGTTCTCTCCGCTGCAGCAAGTTCTATAG 1177

## RESULT 9

ARI144067 ARI144067 150 bp DNA PAT 08-AUG-2001  
LOCUS Sequence 8 from patent US 6210876.  
DEFINITION  
ACCESSION ARI144067  
VERSION ARI144067.1 GI:15105934  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 150)  
AUTHORS Cerney, M.B.  
TITLE Nucleic acid primers and probes for detecting Chlamydia pneumoniae  
JOURNAL Patent: US 6210876-A 8 03-APR-2001;  
FEATURES Location/Qualifiers  
source  
1..150  
/organism="unknown"  
BASE COUNT 43 a 34 c 40 g 33 t  
ORIGIN

Query Match 10.0%; Score 146.8; DB 6; Length 150;

Best Local Similarity 98.7%; Pred. No. 4.2e-25;

Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 366 tacctcaacatcaactagctgacatacaggtgctttgttgagcctccaggatgctgacac 425

Db 1 TACTCTAACATFCATGCTGACATACAGGCTGCTTTGGTGAGCTCCAGGATGCTGTAC 60

Qy 426 taataaagatcacagcgctactgatgagaaacccgaatcgctcgagtgaggaaac 485

Db 61 TAATATAAGGATACAGCGGCTACTGATGAGAAACCGCAATCGCTGGGTGGGAAC 120

Qy 486 taagaatgcgagtgcaattaaagtgggc 515

Db 121 TAAGATGCCGATGCAGTTAAAGTTGGCG 150

## RESULT 10

AX105043 AX105043 816 bp DNA PAT 30-APR-2001  
LOCUS Sequence 19 from Patent WO0125445.  
DEFINITION  
ACCESSION AX105043  
VERSION AX105043.1 GI:13921205  
KEYWORDS  
SOURCE Agkistrodon contortrix.  
ORGANISM Agkistrodon contortrix  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
Viperidae; Crotalinae; Agkistrodon.  
REFERENCE 1 (bases 1 to 816)  
AUTHORS Boone, T.C., Li, H. and Mann, M.B.  
TITLE Fibrinolytically active polypeptide  
JOURNAL Patent: WO 0125445-A 19 12-APR-2001;  
Amgen Inc. (US)  
FEATURES Location/Qualifiers  
source  
1..816  
/organism="Agkistrodon contortrix"  
/db\_xref="taxon:8720"  
/note="Fragment of fibrinolyse of Agkistrodon contortrix"

BASE COUNT 220 a 219 c 168 g 209 t

ORIGIN

Query Match 4.4%; Score 65; DB 6; Length 816;

Best Local Similarity 100.0%; Pred. No. 3.2e-05;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1406 tagaacaacaaactcatctcagaagaggtcgaatagcgccgtcgaccatcatcatc 1465

Db 646 TAGAACAAAACCTCATCTCAGANAGAGGATCTGAATAGGCCGCTCGACCATCATCATC 705



[illegible]







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:50:20 ; Search time 685.35 seconds  
(without alignments)  
1838.868 Million cell updates/sec

Title: US-09-391-606-3  
Perfect score: 1470  
Sequence: 1 atgttaactctattgtcc.....accatcatcatcatcatcat 1470

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1460.4	99.3	2238	21	AA02066		C. pneumoniae 76 k
2	1359.4	92.5	2156	21	AA02063		Chlamydia pneumoniae
3	1348.4	91.7	1230025	20	AA02063		Nucleotide sequenc
4	1347	91.6	1456	21	AA02065		3'-truncated Chlam
5	1155.4	78.6	1852	21	AA02064		5'-truncated Chlam
6	146.8	10.0	150	19	AAV16207		Part of the gene e
7	83	5.6	936	22	AAF58252		Oligonucleotide D1
8	83	5.6	936	22	AAF58254		Oligonucleotide D1
9	83	5.6	936	22	AAF58257		Oligonucleotide D1
10	83	5.6	936	22	AAF58259		Oligonucleotide D2
11	83	5.6	936	22	AAF58262		Oligonucleotide D2

12	83	5.6	938	22	AAF58255		Oligonucleotide D1
13	80.6	5.5	936	22	AAF58252		Oligonucleotide D1
14	80.6	5.5	936	22	AAF58254		Oligonucleotide D1
15	80.6	5.5	936	22	AAF58257		Oligonucleotide D1
16	80.6	5.5	936	22	AAF58259		Oligonucleotide D2
17	80.6	5.5	936	22	AAF58262		Oligonucleotide D2
18	80.6	5.5	938	22	AAF58255		Oligonucleotide D1
19	79.4	5.4	2679	21	AAC60786		Soluble neuropilin
20	75.6	5.1	2670	21	AAC60787		Soluble neuropilin
21	74.8	5.1	2230	22	AAH73017		Human cervical can
22	70	4.8	1550	21	AAA28411		Chlamydia pneumoniae
23	65	4.4	816	22	AA03709		Akistrodon contor
24	65	4.4	1200	22	AA02860		pPIC2ZGPI DNA used
25	65	4.4	1373	22	AA03710		Akistrodon contor
26	65	4.4	1620	22	AA03706		Akistrodon contor
27	65	4.4	1620	22	AA03707		Akistrodon contor
28	63	4.3	1909	21	AAA27110		Rhesus monkey mela
29	63	4.3	4357	22	AAF31389		Expression vector
30	63	4.3	7108	20	AA02662		EP-892047 Seq ID 3
31	63	4.3	7475	20	AA02659		EP-892047 Seq ID 3
32	63	4.3	8192	20	AA02660		EP-892047 Seq ID 3
33	56.4	3.8	72	20	AA029475		p135-p13 construct
34	53.6	3.6	1537	21	AAA63292		C. trachomatis LCV
35	53.6	3.6	1537	22	AAH56196		Chlamydia trachoma
36	50.6	3.4	896	15	AAQ68903		PspA protective an
37	50.6	3.4	946	20	AA033124		Streptococcus pneu
38	50.6	3.4	957	18	AA061726		Streptococcus pneu
39	50.6	3.4	1990	18	AA061725		Streptococcus pneu
40	50.6	3.4	2085	15	AAQ78131		Pneumococcal surfa
41	50.6	3.4	2085	17	AA08979		DNA encoding pneum
42	50.6	3.4	2085	17	AA07178		Pneumococcal surfa
43	50.6	3.4	2085	19	AAV33264		Pneumococcal surfa
44	50.6	3.4	2085	19	AAV39470		Streptococcus pneu
45	50.6	3.4	2085	20	AAZ25063		Streptococcus pneu

## ALIGNMENTS

RESULT 1  
ID AAD02066 standard; DNA; 2238 BP.  
XX AAD02066;  
AC AAD02066;  
DT 26-MAR-2001 (first entry)  
XX C. pneumoniae 76 kDa protein truncation mutant fusion gene.

DE 76 kDa protein; bactericidal; diagnosis; prevention;  
KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;  
KW acute respiratory disease; cough; sore throat; hoarseness; fever;  
KW vaccine; immunisation; treatment; truncation mutant; fusion gene;  
KW ds.

OS Chlamydia pneumoniae.  
OS Synthetic.

FT Key Location/Qualifiers  
FT misc\_feature 1..665 a  
FT /tag=

FT /note= "This part of the sequence is unrelated to  
C. pneumoniae 76 kDa gene"  
FT CDS 766..2238  
FT /tag= b

FT /product= "Truncated Chlamydia pneumoniae  
76kDa protein"  
FT misc\_feature 2122..2238 c  
FT /tag= c

FT /note= "This part of the sequence is unrelated to  
C. pneumoniae 76 kDa gene"  
FT PN WO200065739-A2.



XX 09-NOV-2000.  
 XX 03-MAY-2000; 2000WO-CA00511.  
 XX 03-MAY-1999; 99US-0132270.  
 PR 30-JUN-1999; 99US-0141276.  
 XX (AVET ) AVENTIS PASTEUR LTD.  
 PA Murdin AD, Omen RP, Wang J, Dunn P;  
 PI WPI; 2000-687542/67.  
 XX P-PSDB; AAY71957.  
 DR Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,  
 XX useful for vaccinating against Chlamydia infections -  
 PT Claim 32; Fig 3; 112pp; English.  
 PS The present sequence is a DNA coding for a fusion protein comprising a  
 CC truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded  
 CC residues. C. pneumoniae 76 kDa protein is used in the  
 CC diagnosis, prevention and treatment of C. pneumoniae  
 CC infections (e.g. pneumonia, upper respiratory  
 CC tract disease, bronchitis, sinusitis and acute respiratory  
 CC disease such as cough, sore throat, hoarseness, fever; and  
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence  
 CC is also used as vaccines for immunising humans against diseases  
 CC caused by C. pneumoniae.  
 XX Sequence 2238 BP; 689 A; 459 C; 497 G; 593 T; 0 other;  
 SQ

Query Match 99.3%; Score 1460.4; DB 21; Length 2238;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 1464; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggttaactctattggtccaggtcctatagacgaacagacgacacacctccgcagat 60  
 DB 766 atggttaactctattggtccaggtcctatagacgaacagacgacacacctccgcagat 825  
 QY 61 ctctctgctcaagattgagcgagtcgagcaataaagtcgcggaagctcaagaata 120  
 DB 826 ctctctgctcaagattgagcgagtcgagcaataaagtcgcggaagctcaagaata 885  
 QY 121 gcagggtcggaagctaaagaaatctaaagacgattctgtagagcgatggagatc 180  
 DB 886 gcagggtcggaagctaaagaaatctaaagacgattctgtagagcgatggagatc 945  
 QY 181 ttgcgttctgcagtgatctcagtcctgagcagataagctggggtattgtcttctagt 240  
 DB 946 ttgcgttctgcagtgatctcagtcctgagcagataagctggggtattgtcttctagt 1005  
 QY 241 aacagctgcttctactagcagatctgcagacgtggactcaacacgacgacccgacct 300  
 DB 1006 aacagctgcttctactagcagatctgcagacgtggactcaacacgacgacccgacct 1065  
 QY 301 acgctctctccaccacgtctgattattagacacgacgacgacgacgacgacgacgac 360  
 DB 1066 acgctctctccaccacgtctgattattagacacgacgacgacgacgacgacgacgac 1125  
 QY 361 atctttacctcaacatacactagctgacatacagctgcttggtagcctccaggatgct 420  
 DB 1126 atctttacctcaacatacactagctgacatacagctgcttggtagcctccaggatgct 1185  
 QY 421 gtoactaataaaggatacagcgctactgattgaggaacacgacgacgacgacgacgacgac 480  
 DB 1186 gtoactaataaaggatacagcgctactgattgaggaacacgacgacgacgacgacgacgac 1245  
 QY 481 gaaactaagaatgccgatgcaataaagttggcgcaaatattacagaataatagcgaatat 540  
 DB 1246 gaaactaagaatgccgatgcaataaagttggcgcaaatattacagaataatagcgaatat 1305

QY 541 gcttcggataaaccagcgatttctgactcttttaggttaaacctgacttcttccctcgtcttca 600  
 DB 1306 gcttcggataaaccagcgatttctgactcttttaggttaaacctgacttcttccctcgtcttca 1365  
 QY 601 cagactgctcttctccaaatctgtagcaacaataaacaagcagctgagcttctttaaagag 660  
 DB 1366 cagactgctcttctccaaatctgtagcaacaataaacaagcagctgagcttctttaaagag 1425  
 QY 661 atgcaagataaaccagtagtccacgggaaacgcctgcaattgctcaaatctttagttgat 720  
 DB 1426 atgcaagataaaccagtagtccacgggaaacgcctgcaattgctcaaatctttagttgat 1485  
 QY 721 cagacagatctacagcgacacagatagagaagatggaatgcgattggggatgcataat 780  
 DB 1486 cagacagatctacagcgacacagatagagaagatggaatgcgattggggatgcataat 1545  
 QY 781 ttgcaggacagacgctagtgaggctgtagaaaatgctaaaatctaaataacagataaagc 840  
 DB 1546 ttgcaggacagacgctagtgaggctgtagaaaatgctaaaatctaaataacagataaagc 1605  
 QY 841 aacatagattcagctaaagcagcaaatcgtactgctaaagacacacaaatagctgaagctcag 900  
 DB 1606 aacatagattcagctaaagcagcaaatcgtactgctaaagacacacaaatagctgaagctcag 1665  
 QY 901 aaaaagttcccgactctccaaattctcaagaagcggaacaaatggtaatacagctgag 960  
 DB 1666 aaaaagttcccgactctccaaattctcaagaagcggaacaaatggtaatacagctgag 1725  
 QY 961 aagaatcttaaaaaatacaaaacctgcagatggttctgatgttccaaatccaggaaactaca 1020  
 DB 1726 aagaatcttaaaaaatacaaaacctgcagatggttctgatgttccaaatccaggaaactaca 1785  
 QY 1021 gttgagagctccaaagcaacaaaggaagtagtattggttagtattcgtttccatcgttta 1080  
 DB 1786 gttgagagctccaaagcaacaaaggaagtagtattggttagtattcgtttccatcgttta 1845  
 QY 1081 gatgatgctgaaaaatgagacgcttccattttgatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140  
 DB 1846 gatgatgctgaaaaatgagacgcttccattttgatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1905  
 QY 1141 atgttcaatacggaaaaatcctgattctcaagctcccaacagagagctcgacagcagct 1200  
 DB 1906 atgttcaatacggaaaaatcctgattctcaagctcccaacagagagctcgacagcagct 1965  
 QY 1201 agacagcgaaagccgctggagatgacagctgctgagcgctggcagatgctcagaaaa 1260  
 DB 1966 agacagcgaaagccgctggagatgacagctgctgagcgctggcagatgctcagaaaa 2025  
 QY 1261 gctttagaagcggtcttaggttaaagctgggcaacaaacaggcactactcaatgctttggga 1320  
 DB 2026 gctttagaagcggtcttaggttaaagctgggcaacaaacaggcactactcaatgctttaggga 2085  
 QY 1321 cagactgcttctgctgctgtgtgagcgagagctccctccgctgcagcagagttctatgg 1380  
 DB 2086 cagactgcttctgctgctgtgtgagcgagagctccctccgctgcagcagagttctatgg 2145  
 QY 1381 atccgagctcggttaccagagcttaccgagcaaaaaaactcatctcagaagagatcgtgaat 1440  
 DB 2146 atccgagctcggttaccagagcttaccgagcaaaaaaactcatctcagaagagatcgtgaat 2205  
 QY 1441 agcgcgctgcacatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcat 1470  
 DB 2206 agcgcgctgcacatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcat 2235

RESULT 2  
 AAD02063  
 ID AAD02063 standard; DNA; 2156 BP.  
 XX  
 AC AAD02063;  
 XX  
 DT 26-MAR-2001 (first entry)



XX Chlamydia pneumoniae 76 kDa full-length protein encoding DNA.  
DE 76 kDa protein; bactericidal; diagnosis; prevention; treatment;  
XX pneumonia; upper respiratory tract disease; bronchitis; sinusitis;  
KW acute respiratory disease; cough; sore throat; hoarseness; fever;  
KW vaccine; immunisation; ds.  
XX Chlamydia pneumoniae.  
OS  
XX  
XX Key Location/Qualifiers  
FH 101..2056  
FT /\*tag= a  
FT /product= "Chlamydia pneumoniae 76kDa protein"  
XX  
XX WO200066739-A2.  
XX  
XX 09-NOV-2000.  
XX  
XX 03-MAY-2000; 2000WO-CA00511.  
XX  
XX 03-MAY-1999; 99US-0132270.  
XX 30-JUN-1999; 99US-0141276.  
XX  
XX (AVET ) AVENTIS PASTEUR LTD.  
XX  
XX Murdin AD, Oomen RP, Wang J, Dunn P;  
XX  
XX WPI: 2000-687542/67.  
XX P-PSDB; AAY71954.  
XX  
XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,  
XX useful for vaccinating against Chlamydia infections -  
XX  
XX Claim 2a; Fig 1; 112pp; English.  
XX  
XX The present sequence is a DNA coding for Chlamydia pneumoniae  
XX 76 kDa protein. C. pneumoniae 76 kDa protein  
XX is used in the diagnosis, prevention and treatment  
XX of C. pneumoniae infections (e.g. pneumonia, upper respiratory  
XX tract disease, bronchitis, sinusitis and acute respiratory  
XX disease such as cough, sore throat, hoarseness, fever; and  
XX abnormal chest sounds on auscultation). C. pneumoniae sequence  
XX is also used as vaccines for immunising humans against diseases  
XX caused by C. pneumoniae.  
XX  
XX Sequence 2156 BP; 672 A; 461 C; 471 G; 552 T; 0 other;  
XX

Query Match 92.5%; Score 1359.4; DB 21; Length 2156;  
Best Local Similarity 99.5%; Pred. NO. 0;  
Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
QY 1 atggtaacccattggtccaggtctctatagacgaaacagacacaccccccagat 60  
DB 101 atggtaacccattggtccaggtctctatagacgaaacagacacaccccccagat 160  
QY 61 cttctgctcaaggattggagcgagtgacgaaataagagtgcggaagctcaagaata 120  
DB 161 cttctgctcaaggattggagcgagtgacgaaataagagtgcggaagctcaagaata 220  
QY 121 gcaggtgcgggaagcctaagcctaagcctaagcctaagcctaagcctaagcctaagc 180  
DB 221 gcaggtgcgggaagcctaagcctaagcctaagcctaagcctaagcctaagcctaagc 280  
QY 181 ttgcgttctgcagtgaaagctctctatagctgcgagataagctgggtattctctagt 240  
DB 281 ttgcgttctgcagtgaaagctctctatagctgcgagataagctgggtattctctagt 340  
QY 241 aacagctcgtctctactagcagatctgcagacgtggagctcaacgacgacgcacact 300  
DB 341 aacagctcgtctctactagcagatctgcagacgtggagctcaacgacgacgcacact 400

QY 301 acgctctccaccacccacgtctgtattataagactcaagcgaaacagcttacgatact 360  
DB 401 acgctctccaccacccacgtctgtattataagactcaagcgaaacagcttacgatact 460  
QY 361 acctttaccctaacatcactactagctacatacaggtgctttgtgagctccagatgct 420  
DB 461 atctttactctaacatcactactagctacatacaggtgctttgtgagctccagatgct 520  
QY 421 gtccactaatataaaggatacagcgctactgatgaggagaaacgcaatcgctcgagtg 480  
DB 521 gtccactaatataaaggatacagcgctactgatgaggagaaacgcaatcgctcgagtg 580  
QY 481 gaaactaagaatgcgcgagcgaattaaagtggcgcgcaaatctacagaattagcgaatat 540  
DB 581 gaaactaagaatgcgcgagcgaattaaagtggcgcgcaaatctacagaattagcgaatat 640  
QY 541 gcttcggataaaccagcgattcttgactcttttaggttaaaactgactctcttcgacctta 600  
DB 641 gcttcggataaaccagcgattcttgactcttttaggttaaaactgactctcttcgacctta 700  
QY 601 cagactgctctctccaatctgtagcaacaataaacaagcagctgagctctttaaagag 660  
DB 701 cagggctgctctctccaatctgtagcaacaataaacaagcagctgagctctttaaagag 760  
QY 661 atgcaagataaccacgtagtcccgagggaacgcctgcaattgctcaattcttagttgat 720  
DB 761 atgcaagataaccacgtagtcccgagggaacgcctgcaattgctcaattcttagttgat 820  
QY 721 cagacagatgctacagcgacacagatagagaagaatggaatgcgattggggatgcata 780  
DB 821 cagacagatgctacagcgacacagatagagaagaatggaatgcgattggggatgcata 880  
QY 781 ttgcagacagacagctagtgagctgtagaaatgctaaatctaaatcaatacagataagc 840  
DB 881 ttgcagacagacagctagtgagctgtagaaatgctaaatctaaatcaatacagataagc 940  
QY 841 aacatagattcagctaaagcagcaatcgctactgctaaagacacaaatagctgaagctcag 900  
DB 941 aacatagattcagctaaagcagcaatcgctactgctaaagacacaaatagctgaagctcag 1000  
QY 901 aaaaagtctcccgactctccaatcttcaagaagcggaacaaatggtaacagctgag 960  
DB 1001 aaaaagtctcccgactctccaatcttcaagaagcggaacaaatggtaacagctgag 1060  
QY 961 aagatcttaaaaaataatacaaacctgcagatgctctgatttccaaatccaggaactaca 1020  
DB 1061 aagatcttaaaaaataatacaaacctgcagatgctctgatttccaaatccaggaactaca 1120  
QY 1021 gttggaggtcccaagcaacaaaggaagtagtattggttagtattggtttccatgcttta 1080  
DB 1121 gttggaggtcccaagcaacaaaggaagtagtattggttagtattggtttccatgcttta 1180  
QY 1081 gatgatgtaaaatgagaccgcttccatttgcatttgcatttgcatttgcatttgcatttgc 1140  
DB 1181 gatgatgtaaaatgagaccgcttccatttgcatttgcatttgcatttgcatttgcatttgc 1240  
QY 1141 atgttcaatacaggaataatcctgattctcaagctgcccacagagagctcgacacaaagct 1200  
DB 1241 atgttcaatacaggaataatcctgattctcaagctgcccacagagagctcgacacaaagct 1300  
QY 1201 agagcaggaagcgcgctggagatgacagtgctgctgcagcgtggcagatgctcagaaa 1260  
DB 1301 agagcaggaagcgcgctggagatgacagtgctgctgcagcgtggcagatgctcagaaa 1360  
QY 1261 gctttagaagcggctcttaggttaaaagctgggcaacaaagggcagataactcaatgctttggga 1320  
DB 1361 gctttagaagcggctcttaggttaaaagctgggcaacaaagggcagataactcaatgctttggga 1420  
QY 1321 cagatcgtcttctgctgctgtgtgagcgaggag-ccctcccgctgcagcagagcttctatg 1379  
DB 1421 cagatcgtcttctgctgctgtgtgagcgaggagttctcccgctgcagcagagcttctatg 1480  
QY 1380 g 1380



Db	1481 g. 1481	
RESULT	3	
AA91990/c		
ID	AA91990 standard; DNA; 1230025 BP.	
XX		
AC	AA91990;	
XX		
DT	13-SEP-1999 (first entry)	
XX		
DE	Nucleotide sequence of the complete genome of Chlamydia pneumoniae.	
XX		
KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;	
KW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;	
KW	vaccine; neutralising epitope; ss.	
XX		
OS	Chlamydia pneumoniae.	
XX		
PN	W09927105-A2.	
XX		
PD	03-JUN-1999.	
XX		
XX	20-NOV-1998; 98WO-IB01890.	
PF		
XX		
XX	04-NOV-1998; 98US-0107078.	
PR		
PR	21-NOV-1997; 97FR-0014673.	
XX		
XX	(GEST ) GENSET.	
XX		
PI	Griffais R;	
XX		
XX	WPI; 1999-357842/30.	
DR		
XX	Genome sequence of Chlamydia pneumoniae	
PT		
XX	Claim 1; Page 291-611; 1912pp; English.	
PS		
XX		
CC	The present sequence represents the complete genome of Chlamydia	
CC	pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes	
CC	respiratory disease such as pneumonia and bronchitis and is thought	
CC	to be a contributing factor in heart disease, sarcoidosis, sinusitis,	
CC	purulent otitis media, erythema nodosum or pharyngitis. The polypeptides	
CC	encoded by the open reading frames of the C. pneumoniae genome (see	
CC	AA91990) can be used in immunogenic compositions as vaccines.	
CC	Vectors containing C. pneumoniae nucleotide sequences can also be	
CC	used as immunogenic compositions, especially where the vector directs	
CC	the expression of a neutralising epitope of C. pneumoniae.	
XX		
SQ	Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;	
Query Match	91.7%; Score 1348.4; DB 20; Length 1230025;	
Best Local Similarity	99.4%; Pred. No. 0;	
Matches 1374; Conservative	0; Mismatches 6; Indels 2; Gaps 2;	
QY	1 atgtttaatcctattgtccaggtcctatagacgaacagacgacaccccgagat 60	
DB	828598 ATGTTTAAATCCTATTGTCAGGTCCTATAGACGAACAGACACACCTCCCGCAGAT 828539	
QY	61 ctttctgtcgaagattggagggcgtgcagcaataaagagtcgggaagctcaagaata 120	
DB	828538 CTTTCTGCTCAAGGATTGGAGCGGAGTGCAGCAATTAAGAGTGGGAAGCTCAAGAATA 828479	
QY	121 gcaggtcggaagctaaagcctaaagcctaaagcctaaagcctaaagcctaaagcct 180	
DB	828478 GCAGTGGCGGAAGCTAAAGCCTAAAGATCTAAGACCAACAGACACCTCCCGCAGAT 828419	
QY	181 ttgcgttctcagtgaaatgctctcctagctgagtcgacataagctgggtattgctctagt 240	
DB	828418 TTGCGTTCGAGTAATGCTCTCATGAGTTCGCAGATAAGTGGGTATTGCTTCTAGT 828359	

QY	241	aacagctcgtcttctactagcagatctgcagacgtgagctcaacacgacgaccccaact 300
DB	828358	AACAGCTCGTCTTCTACTAGCAGATCTGACAGCTGGACTCAACGACGACGACACCT 828299
QY	301	acgctcctcccccacccgctgtgattataagactcaagcgcacaaacagcttacatact 360
DB	828298	ACGCTCTCTCCACCCACGTTGATGATTATAAGACTCAAGCGCAACACAGCTTACGATAT 828239
QY	361	atctttacctcaacatcactagctacacagctgcttctggtgagctccagatgct 420
DB	828238	ATCTTTACCTCAACATCAGCTAGCTGACATACAGGCTGCTTTGCTGAGCCTCCAGGATGCT 828179
QY	421	gtcacataataaagatacagcggtactgatgaggaacccgcaatcgctgcggagtgg 480
DB	828178	GTCACATAATAAAGGATACAGCGGTACTGATGAGAAACCGCAATCGTCGGAGATGG 828119
QY	481	gaactaagaatccgcatgcaataaagtggcgcaaatcacgaattacgaataat 540
DB	828118	GAACCTAAGAAATCCGATGAGTTAAGTTGGCGCGCAATTAACGAATTAAGCAATAT 828059
QY	541	gcttcgataaccaagcgtattctgactcttttagttaaactgacttcttgcaccttta 600
DB	828058	GCTTCGGATAACCAAGGATCTTCTGACTCTTTAGGTAAGTACTCTTCGACCTCTTA 827999
QY	601	cagactgctcttctccaatctgtagcaacaataaacaagcagctgagcttctaaag 660
DB	827998	CAGGCTGCTCTCTCCCAATCTGTAGCAACAAATAAACAAAGCAGCTGAGTCTTAAAG 827939
QY	661	atcaagataaccagtagtcccgaggaacccgctcaatctgctcaatctttagttgat 720
DB	827938	ATGCAAGATAACCCAGTAGTCCCGAGGAACCGCTGCATTTGCTCAATCTTTAGTTGAT 827879
QY	721	cagacagatgctacagcgacacagatagagaagatggaatgcgattggggatgcata 780
DB	827878	CAGACAGATGCTACAGCGACACAGATAGAGAAAGATGGAATGCGATTAGGATGCATAT 827819
QY	781	tttcaggaacagacgctagtgagctgtgagaaatctcaaatctaaacagctataagc 840
DB	827818	TTTCAGGACAGAACGCTAGTGGAGCTGTAGAAATGCTAAATCTAATACAGTAATAGC 827759
QY	841	aactagattcagctaaagcagcaatcgctactgtcaagacacaaatagctgaagctcag 900
DB	827758	AACATAGATTTCAGCTTAAGCAGCAATCGCTACTGCTAAGACACAAATAGCTGAAGCTCAG 827699
QY	901	aaaagtctcccgactctccatcttcaagaagcgaacaaatggtaatacagctcag 960
DB	827698	AAAAAGTTCCCGACTCTCCAATTTCTCAAGAAGCGGAACAAATGGTAATACAGGCTGAG 827639
QY	961	aaagatcttaaaatatacaacctgcagatggttctgatgttccaaatcccggaactaca 1020
DB	827638	AAAGATCTTAAATAATATCAAACTGCAGATGTTCTGATGTTCCAAATCCAGGAATACA 827579
QY	1021	gttgaggtctcaagcaacaaaggaagtagtattggttagtattctgtttccatgcgtta 1080
DB	827578	GTTGGAGGCTCCCAAGCAACAGGAAGTAGTATTGGTAGTATTTCGTGTTTCCATGCTGTTA 827519
QY	1081	gatgatgtgaaatgagacgcttccattttgatgtctggttttcgcatgattcac 1140
DB	827518	GATGATGCTGAAATAGAGACCCCTTCCATTTTGTGATGCTGGGTTTCGTCAGATGATTCAC 827459
QY	1141	atgttcaatcacgaaatcctgattctcaagctgcccaacagagctgcgacgacaaagct 1200
DB	827458	ATGTTCAATACCGAAATCCTGATCTCAAGTGCCTCAACAGAGAGCTCGACGACAAAGCT 827399
QY	1201	agagcagcgaagcgcgtggagatgacagtgctgctgcagcgtgcgagatgctcagaaa 1260
DB	827398	AGAGCAGCGAAAGCCGCTGGAGATGACAGTGTGCTGAGCGCTGGCAGATGCTCAGAAA 827339
QY	1261	gctttagaagcggctcagtaagctggaacacagcggcagcagcagcagcagcagcagc 1320
DB	827338	GCTTTAGAAGCGGCTCTAGGTAAAGCTGGGCAACAAAGGAGCTACTCAATGCTTTAGGA 827279
QY	1321	cagatcgcttctgctgctg-ttgtgagcgaggag-tcctcccgctgcagcaagttctat 1378











Db 901 tccatttgatgctggtttctgctcagatgattcacatgtttcaatacaggaatacctctgat 960  
Qy 1165 tctcaagctgcccacacagagctcgacacaaagctagacagcagcagaaagccgctggagat 1224  
Db 961 tctcaagctgcccacacagagctcgacacaaagctagacagcagcagaaagccgctggagat 1020  
Qy 1225 gacagtgctgctgcagcgcgtgcagatgctcagaaaagcttttagaagcgcctctaggtaaa 1284  
Db 1021 gacagtgctgctgcagcgcgtgcagatgctcagaaaagcttttagaagcgcctctaggtaaa 1080  
Qy 1285 gctgggcaacacagggcactactcaatgctttggacagatcgcttctgctgctgtgtg 1344  
Db 1081 gctgggcaacacagggcactactcaatgcttttagaagcagatcgcttctgctgctgtgtg 1140  
Qy 1345 agcgcagag-tcctcccgctgcagcaagtctatgg 1380  
Db 1141 agcgcagagttctcccgctgcagcaagtctatag 1177

## RESULT 6

AAV16207  
ID AAV16207 standard; DNA; 150 BP.

XX AAV16207;

XX 28-MAY-1998 (first entry)

XX Part of the gene encoding the 76 kDa cystein rich OMP.

XX Cysteine rich outer-membrane protein; OMP; 60 kDa OMP; 76 kDa OMP;  
XX hybridisation; amplification; assay; detection; ds.

XX Chlamydia pneumoniae.

XX WO9746709-A2.

XX 11-DEC-1997.

XX 04-JUN-1997; 97WO-US09673.

XX 06-JUN-1996; 96US-0659473.

XX (ABBO ) ABBOTT LAB.

XX Cerney MB;

XX WPI; 1998-042215/04.

XX Chlamydia pneumoniae derived oligonucleotides - used as primers or  
XX probes for specific and sensitive detection

XX Disclosure; Page 24; 28pp; English.

XX The present sequence represents part of the gene encoding the 76 kDa  
XX cysteine rich outer-membrane protein (OMP) of Chlamydia pneumoniae.  
XX Novel oligonucleotides AAV16195-206 can function either as PCR primers  
XX or probes. They are used to detect C. pneumoniae in a sample.  
XX Oligonucleotides V161202-04 are specific for the present sequence.  
XX The oligonucleotides can be used in a hybridisation or amplification  
XX based assay for the detection of C. pneumoniae in a test sample.

XX Sequence 150 BP; 43 A; 34 C; 40 G; 33 T; 0 other;

## Query Match

Best Local Similarity 10.08; Score 146.8; DB 19; Length 150;  
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 366 tacctcaacatactagctgacatacagctgcttttggtagcctccaggatgctgtcac 425

Db 1 tacctcaacatactagctgacatacagctgcttttggtagcctccaggatgctgtcac 60

Qy 426 taataaagagatacagcgcgtactgtaggaaacgcgcaatcgctgcggagtgggaaac 485

Db 61 taataaagagatacagcgcgtactgtaggaaacgcgcaatcgctgcggagtgggaaac 120  
Qy 486 taagaatgccgatgcaatataaagttggcgc 515  
Db 121 taagaatgccgatgcaatataaagttggcgc 150

## RESULT 7

AAF58252/c  
ID AAF58252 standard; DNA; 936 BP.

XX AAF58252;

XX 24-APR-2001 (first entry)

XX Oligonucleotide D1835.

XX Electron-transfer group; ETM; mismatch; genotyping;

XX gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in  
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on  
XX a single surface

XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic  
XX acids each containing an electron-transfer group (ETM) having  
XX different redox potentials. The invention is used for electronic  
XX detection of nucleic acids, especially of substitutions (mismatches)  
XX and single-nucleotide polymorphisms, e.g. for genotyping,  
XX monitoring gene expression.

XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

## Query Match

Best Local Similarity 5.6%; Score 83; DB 22; Length 936;  
Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

Qy 482 aaactaagaatgcgcgatgcaatataaagttggcgcgcaatatacagaatagcgaatag 541

Db 750 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 691

Qy 542 ctctggataacacagcgattcttgactcttttagttaaactgacttcttcgacctttac 601

Db 690 GWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 631

Qy 602 agactgctcttctccaatctgtagcaacaatacaaacagcagctgagcttctttaagaga 661

Db 630 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 571

Qy 662 tgcagaataacccagtagtccccagggaaacgcgcgaatgctgaatctttagttgatac 721

Db 570 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 511







OS Synthetic.  
 PN WO200107665-A2.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 26-JUL-2000; 2000WO-US20476.  
 XX  
 PR 26-JUL-1999; 99US-0145695.  
 PR 17-MAR-2000; 2000US-0190259.  
 XX  
 PA (CLIN-) CLINICAL MICRO SENSORS INC.  
 XX  
 PI Umek RM;  
 XX  
 DR WPI; 2001-159728/16.  
 XX  
 PT Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX  
 PS Example 6; Page 127; 159pp; English.  
 XX  
 CC The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX  
 SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 5.6%; Score 83; DB 22; Length 936;  
 Best Local Similarity 1.2%; Pred. No. 7.9e-14;  
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

QY 482 aaactaagaatgccgatgcaattaaagtgtggtgcgcgaattacagaattagcgaataatg 541  
 DB 750 WWWWWW  
 QY 542 ctccggataaccagcagattcttgactcttttagttaaactgactcttccttcac 601  
 DB 690 GWWWWW  
 QY 602 agactgtctcttcccaatctgtagcaacaataacaaagcagctgagctctttaaagaga 661  
 DB 630 WWWWWW  
 QY 662 tgaagataaccagtagtccaggggaaacgcctgcaattgctcaattcttagtgatc 721  
 DB 570 WWWWWW  
 QY 722 agacagatgctacgcagcacagatagagaagatggaatgcgattgggagatcatatt 781  
 DB 510 WWWWWW  
 QY 782 ttgcaggacagacgcagtagtgagctgtagaaaatgtcaaatcttaataacagtataagca 841  
 DB 450 WWWWWW  
 QY 842 acatagattcagctaaagcagcaatcgctactgtactgtactgaacacaaatagctgaagctcaga 901  
 DB 390 WWWWWW  
 QY 902 aaaagtcccccactctcaattcttcaagaagcgaacaaatgtaataacagctgaga 961  
 DB 330 WWWWWW  
 QY 962 agagatttaaaaaatatacaaacctgcagatggttctgtatgttcccaatccaggaactacag 1021  
 DB 270 WWWWWW

QY 1022 ttggagggtcccaagcaacaagaagtagtattgttgtagtattgtttccatgctgttag 1081  
 DB 210 WWWWWW  
 QY 1082 atgagtgtgaaaatgagaccgcttccatttgcattgtctgggttcgtcagatgattcaca 1141  
 DB 150 WWWWWW  
 QY 1142 tggtaataacggaataatctcgtattctcaa 1170  
 DB 90 WWWWWW

RESULT 10  
 AAF58259/c  
 ID AAF58259 standard; DNA; 936 BP.  
 XX  
 AC AAF58259;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Oligonucleotide D2004.  
 XX  
 KW Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 OS Synthetic.  
 PN WO200107665-A2.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 26-JUL-2000; 2000WO-US20476.  
 XX  
 PR 26-JUL-1999; 99US-0145695.  
 PR 17-MAR-2000; 2000US-0190259.  
 XX  
 PA (CLIN-) CLINICAL MICRO SENSORS INC.  
 XX  
 PI Umek RM;  
 XX  
 DR WPI; 2001-159728/16.  
 XX  
 PT Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX  
 PS Example 6; Page 128; 159pp; English.  
 XX  
 CC The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX  
 SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 5.6%; Score 83; DB 22; Length 936;  
 Best Local Similarity 1.2%; Pred. No. 7.9e-14;  
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

QY 482 aaactaagaatgccgatgcaattaaagtgtggtgcgcgaattacagaattagcgaataatg 541  
 DB 750 WWWWWW  
 QY 542 ctccggataaccagcagattcttgactcttttagttaaactgactcttccttcac 601  
 DB 690 GWWWWW  
 QY 602 agactgtctcttcccaatctgtagcaacaataacaaagcagctgagctctttaaagaga 661  
 DB 630 WWWWWW











XX Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface -  
XX  
XX Example 6; Page 127; 159pp; English.  
PS  
XX The present invention relates to a composition comprising two nucleic  
XX acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
CC  
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;  
XX  
SQ

[illegible]

RESULT 15  
AAF58257  
ID AAF58



XX	AAF58257;
AC	
XX	24-APR-2001 (first entry)
DT	
XX	Oligonucleotide D1954.
DE	
XX	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
KW	
XX	Synthetic.
OS	
XX	WO200107665-A2.
PN	
XX	01-FEB-2001.
PD	
XX	26-JUL-2000; 2000WO-US20476.
XX	
PF	26-JUL-1999; 990US-0145695.
XX	
PR	17-MAR-2000; 2000US-0190259.
PR	
XX	(CLIN-) CLINICAL MICRO SENSORS INC.
PA	
XX	Umek RM;
PI	
XX	WPI; 2001-159728/16.
DR	
XX	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface
PT	
XX	Example 6; Page 127; 159pp; English.
PS	
XX	The present invention relates to a composition comprising two nucleic
CC	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
SQ	Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

```

Query Match      5.5%; Score 80.6; DB 22; Length 936;
Best Local Similarity 1.2%; Pred. No. 3.9e-13;
Matches      8; Conservative 402; Mismatches 281; Indels      0; Gaps

```

QY	483	aactaagaatgcogatgcattaaagtggcgcgaattacagaaattacagaaattacgcaaatatgc	542
Db	1	#####	60
QY	543	ttcggataaccaacgactcttgactctttaggtaaaactgacttcttcgcacctcttaca	602
Db	61	#####	120
QY	603	gactgctcttctccaatctgtagcaacaataacaagcagctgagctctttaaagagat	662
Db	121	#####	180
QY	663	gcaagataaccacgtagtcccagggaacgctgcaattgctcaatctttagtgatca	722
Db	181	#####	240
QY	723	gacagatgctacagcacagatagagaaagatggaatgcgattgggagtcataatt	782
Db	241	#####	300
QY	783	tgcaaggacagaacgctagtggagctgtagaaaatgctaaatcaataacagtatagaacaa	842
Db	301	#####	360
QY	843	catagattcagctaaagcagcaatcgctactgtcgtactgaagcacaaaatagctgaagctcga	902
		#####	

Search completed: February 7, 2002, 19:59:45  
Job time: 15099 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:53:37 ; Search time 243.49 Seconds  
(without alignments)  
1367.295 Million cell updates/sec

Title: US-09-391-606-3

Perfect score: 1470

Sequence: 1 atggttaatctctattgttc.....accatcatcatcatcat 1470

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146.8	10.0	150	4	US-08-659-473-8
2	65	4.4	816	4	US-09-411-329C-19
3	65	4.4	1373	4	US-09-411-329C-20
4	65	4.4	1620	4	US-09-411-329C-12
5	65	4.4	1620	4	US-09-411-329C-13
6	54.4	3.7	7218	1	US-08-232-463-14
7	50.6	3.4	125	2	US-08-350-260A-41
8	50.6	3.4	2085	1	US-08-072-070-1
9	50.6	3.4	2085	1	US-08-465-746-1
10	50.6	3.4	2085	1	US-08-214-184-1
11	50.6	3.4	2085	1	US-08-469-434-1
12	50.6	3.4	2085	1	US-08-214-222-1
13	50.6	3.4	2085	2	US-08-467-852A-1
14	50.6	3.4	2085	2	US-08-468-718-1
15	50.6	3.4	2085	2	US-08-246-636-1
16	50.6	3.4	2085	2	US-08-247-431A-1
17	50.6	3.4	2085	2	US-08-319-795-1
18	50.6	3.4	2085	2	US-08-468-985-1
19	50.6	3.4	2086	3	US-08-312-949-1
20	50.6	3.4	2086	3	US-08-446-201-2
21	50.4	3.4	217	2	US-08-958-201-17
22	49.8	3.4	913	2	US-08-661-052-13
23	49.8	3.4	913	4	US-09-188-082-13
24	49.8	3.4	1679	2	US-08-661-052-15
25	49.8	3.4	1679	4	US-09-188-082-15
26	49	3.3	907	3	US-09-184-658-9
27	48.8	3.3	216	2	US-08-652-816A-40

28 43.6 3.0 6744 1 US-08-119-125A-2 Sequence 2, Appli  
29 39.8 2.7 72 2 US-08-273-146-36 Sequence 36, Appl  
30 39.8 2.7 123 2 US-08-273-146-38 Sequence 38, Appl  
31 39.8 2.7 132 2 US-08-273-146-40 Sequence 40, Appl  
32 39.8 2.7 274 2 US-08-448-418-96 Sequence 96, Appl  
33 39.8 2.7 883 3 US-09-184-658-7 Sequence 7, Appli  
34 39.6 2.7 126 1 US-08-211-202-131 Sequence 131, App  
35 39.6 2.7 162 1 US-08-211-202-132 Sequence 132, App  
36 39.6 2.7 162 2 US-08-448-418-94 Sequence 94, Appl  
37 39.6 2.7 640 3 US-08-961-083-33 Sequence 33, Appl  
38 39.6 2.7 4403 5 PCT-US93-01901-30 Sequence 30, Appl  
39 39.4 2.7 840 4 US-09-260-527-4 Sequence 4, Appli  
40 39.4 2.7 3337 1 US-08-072-610-1 Sequence 1, Appli  
41 39.4 2.7 3337 2 US-08-719-822B-1 Sequence 1, Appli  
42 39.4 2.7 3337 4 US-09-092-458-1 Sequence 1, Appli  
43 38 2.6 6896 2 US-08-627-151A-6 Sequence 6, Appli  
44 37.4 2.5 1920 1 US-08-186-222-1 Sequence 1, Appli  
45 37.4 2.5 5361 4 US-08-973-462-2 Sequence 2, Appli

#### ALIGNMENTS

RESULT 1  
US-08-659-473-8  
; Sequence 8, Application US/08659473  
; Patent No. 6210876  
; GENERAL INFORMATION:  
; APPLICANT: M. B. Cerney  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.0.1  
; SOFTWARE: Microsoft Word 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/659,473  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul D. Yasger  
; REGISTRATION NUMBER: 37,477  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 708/937-2341  
; TELEFAX: 708/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 150 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA (C. pneumoniae)  
US-08-659-473-8

Query Match 10.0%; Score 146.8; DB 4; Length 150;  
Best Local Similarity 98.7%; Pred. No. 3.2e-34;  
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 366 tacctcaacatcactagctgacatcacagctcttttgtagcctccagatgctgtcac 425  
|||||  
Db 1 TACCTCAACATCAGCTAGCTGACATACAGGCTGCTTTGGTGACCTCCAGATGCTGTAC 60



Qy 426 taatataaagatagagcggtactatgaggaagaaacgcgcgtcgtcagtgaggaaac 485  
 Db 61 TAATATAAGGATACAGCGGCTACTGATGAGGAACCCGCAATCGCTCGGTGTGGGAAC 120  
 Qy 486 taagaatccgcatgaattaaagtggcgc 515  
 Db 121 TAAGATCCGATGCGTAAAGTTGGCGC 150

RESULT 2  
 US-09-411-329C-19  
 ; Sequence 19, Application US/09411329C  
 ; Patent No. 6261820  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boone, Thomas  
 ; APPLICANT: Li, Huimin  
 ; APPLICANT: Mann, Michael  
 ; TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE  
 ; FILE REFERENCE: A-596  
 ; CURRENT APPLICATION NUMBER: US/09/411,329C  
 ; CURRENT FILING DATE: 1999-10-01  
 ; NUMBER OF SEQ ID NOS: 29  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 19  
 ; LENGTH: 816  
 ; TYPE: DNA  
 ; ORGANISM: Agkistrodon contortrix  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(816)  
 ; OTHER INFORMATION: Fragment of fibrolase of Agkistrodon contortrix  
 US-09-411-329C-19

Query Match 4.4%; Score 65; DB 4; Length 816;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1406 tagaacaataactctcagaagagatctgaatagcgccgtcgaccatcatcatc 1465  
 Db 646 tagaacaataactctcagaagagatctgaatagcgccgtcgaccatcatcatc 705  
 Qy 1466 atcat 1470  
 Db 706 atcat 710

RESULT 3  
 US-09-411-329C-20  
 ; Sequence 20, Application US/09411329C  
 ; Patent No. 6261820  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boone, Thomas  
 ; APPLICANT: Li, Huimin  
 ; APPLICANT: Mann, Michael  
 ; TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE  
 ; FILE REFERENCE: A-596  
 ; CURRENT APPLICATION NUMBER: US/09/411,329C  
 ; CURRENT FILING DATE: 1999-10-01  
 ; NUMBER OF SEQ ID NOS: 29  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 20  
 ; LENGTH: 1373  
 ; TYPE: DNA  
 ; ORGANISM: Agkistrodon contortrix  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(1373)  
 ; OTHER INFORMATION: Fragment of fibrolase of Agkistrodon contortrix  
 US-09-411-329C-20

Query Match 4.4%; Score 65; DB 4; Length 1373;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-09;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1406 tagaacaataactctcagaagagatctgaatagcgccgtcgaccatcatcatc 1465  
 Db 1203 tagaacaataactctcagaagagatctgaatagcgccgtcgaccatcatcatc 1262  
 Qy 1466 atcat 1470  
 Db 1263 atcat 1267

RESULT 4  
 US-09-411-329C-12  
 ; Sequence 12, Application US/09411329C  
 ; Patent No. 6261820  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boone, Thomas  
 ; APPLICANT: Li, Huimin  
 ; APPLICANT: Mann, Michael  
 ; TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE  
 ; FILE REFERENCE: A-596  
 ; CURRENT APPLICATION NUMBER: US/09/411,329C  
 ; CURRENT FILING DATE: 1999-10-01  
 ; NUMBER OF SEQ ID NOS: 29  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 12  
 ; LENGTH: 1620  
 ; TYPE: DNA  
 ; ORGANISM: Agkistrodon contortrix  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(1620)  
 ; OTHER INFORMATION: Complementary (sense) strand of antisense strand (See SEQ ID  
 ; OTHER INFORMATION: 3  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(1620)  
 ; OTHER INFORMATION: Coding sequence of native pro-fibrolase of Agkistrodon contor  
 US-09-411-329C-12

Query Match 4.4%; Score 65; DB 4; Length 1620;  
 Best Local Similarity 100.0%; Pred. No. 2e-09;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1406 tagaacaataactctcagaagagatctgaatagcgccgtcgaccatcatcatc 1465  
 Db 1412 tagaacaataactctcagaagagatctgaatagcgccgtcgaccatcatcatc 1471  
 Qy 1466 atcat 1470  
 Db 1472 atcat 1476

RESULT 5  
 US-09-411-329C-13/c  
 ; Sequence 13, Application US/09411329C  
 ; Patent No. 6261820  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boone, Thomas  
 ; APPLICANT: Li, Huimin  
 ; APPLICANT: Mann, Michael  
 ; TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE  
 ; FILE REFERENCE: A-596  
 ; CURRENT APPLICATION NUMBER: US/09/411,329C  
 ; CURRENT FILING DATE: 1999-10-01  
 ; NUMBER OF SEQ ID NOS: 29  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 13  
 ; LENGTH: 1620  
 ; TYPE: DNA  
 ; ORGANISM: Agkistrodon contortrix  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(1620)  
 ; OTHER INFORMATION: Complementary (antisense) strand of sense strand (See SEQ ID



OTHER INFORMATION: 2  
NAME/KEY: misc.feature  
OTHER INFORMATION: Anti-coding sequence of native pro-fibrolase of Agkistrodon contortrix  
OTHER INFORMATION: rtri  
US-09-411-329C-13

Query Match 4.4%; Score 65; DB 4; Length 1620;  
Best Local Similarity 100.0%; Pred. No. 2e-09;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1406 tagaacaacacatctcagaagaggtctgaatagcgcctcagccatcatcatc 1465  
DB 209 TAGAACAAACATCTCTCAGAGAGGATCTGAATAGCCCGCTGACCATCATCATC 150  
QY 1466 atcat 1470  
DB 149 ATCAT 145

RESULT 6  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232.463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935.313  
FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 3.7%; Score 54.4; DB 1; Length 7218;  
Best Local Similarity 9.3%; Pred. No. 6.6e-06;  
Matches 43; Conservative 218; Mismatches 199; Indels 0; Gaps 0;

QY 631 aataaacaagcagctgagcttctttaaagagatgaagaataaaccagtagtccacaggaaa 690  
DB 1481 AATTACCTATCTATCAAGTAGTTAAAGAGATAGAGAATTTGGTACRRRRRRRRRRR 1422  
QY 691 acgcctgaattgctcaatctttagtgatcagacagatgctacagcagacagatagag 750  
DB 1421 RFRRR 1362  
QY 751 aaagatgaaatcgatgggagatcatatttgcaggacagaaacgctagtgagctgta 810  
DB 1361 RFRRR 1302  
QY 811 gaaatgctcaatctaaacagtaaaagcaacacagatcagctaaagcagcaaatcgct 870  
DB 1301 RFRRR 1242  
QY 871 actgtaagacacaaatagctgaagctcagaaaaagttcccgactcccaatttcaa 930  
DB 1241 RFRRR 1182  
QY 931 gaagcgaacaaatgtaatacagctgagaaagatcttaaaaaatatacaaacctcagat 990  
DB 1181 RFRRR 1122  
QY 991 ggtctgagttccaaatccaggaactacagttgaggtcctcaacaaagaaagtagt 1050  
DB 1121 RFRRR 1062  
QY 1051 attgtagtattcggttccatgctgtagatgtagtctg 1090  
DB 1061 AAGTCCCTCGACCTGCAGCCAAAGCTCGGAATTAATTCG 1022

RESULT 7  
US-08-350-260A-41  
Sequence 41, Application US/08350260A  
Patent No. 5962255  
GENERAL INFORMATION:  
APPLICANT: Winter, Gregory Paul  
APPLICANT: Griffiths, Andrew David  
APPLICANT: Williams, Samuel Cameron  
APPLICANT: Waterhouse, Peter  
APPLICANT: Nissim, Ahuva  
APPLICANT: Johnson, Kevin Stuart  
APPLICANT: Smith, Andrew John Hammond  
TITLE OF INVENTION: Methods for producing members of specific  
TITLE OF INVENTION: binding pairs  
NUMBER OF SEQUENCES: 602  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David W. Clough  
STREET: Marshall, O'Toole, Gerstein, Murray & Borun  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350.260A  
FILING DATE: 05-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9110549.4  
FILING DATE: 15-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206318.9  
FILING DATE: 24-MAR-1992  
PRIOR APPLICATION DATA:



```

RESULT          9
US-08-465-746-1
; Sequence 1, Application US/08465746
; Patent No. 5679768
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: Vother, Janet L
; APPLICANT: McDaniel, Larry S
; TITLE OF INVENTION: EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEIN A
; NUMBER OF SEQUENCES: 5

```



CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Mattare, Ltd  
 STREET: Suite 1203, 2001 Jefferson Davis Highway  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 FILING DATE: 12-FEB-1992  
 APPLICATION NUMBER: US/08/465,746  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/048,896  
 FILING DATE: 15-FEB-1991  
 APPLICATION NUMBER: US 07/656,773  
 FILING DATE: 12-FEB-1992  
 APPLICATION NUMBER: US 07/835,698  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 521-0378  
 TELEX: LURPAT WASHINGTON  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2085 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Streptococcus pneumoniae  
 STRAIN: Rx1  
 IMMEDIATE SOURCE:  
 CLONE: JY4313  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 1..2085  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: join(127..1984)  
 US-08-465-746-1

Query Match 3.4%; Score 50.6; DB 1; Length 2085;  
 Best Local Similarity 49.8%; Pred. No. 4.3e-05;  
 Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;  
 QY 745 atagagaagaatggaatgcgattgggagatgcatatttgcaggacagaaagcgtagtga 804  
 DB 472 ACAGACAAGCGCGAAGAACGACGACGACGACGACGACGACGACGACGACGACGACG 531  
 QY 805 gctgtagaaaatgctaaatcctaataacagataaagcagataaagcagataaagcagca 864  
 DB 532 GAAGAGGCAAAACTAAATTTAACTGTTTCGAGCAATGGTAGTTCCTGAGCCAGCAG 591  
 QY 865 atcgtactgctaagacacaaaatagctgaagctcagaaaaagttcccca 921  
 DB 592 TTGGCTGAGACTAAGAAAAATCAGACAGCTAAACAAAAGCACCAGACTTACTTAA 651  
 QY 922 attctcagaagcggcaaatggaatgtaacagctgaggaagatcttataaatcaaa 981  
 DB 652 AAACATAGAGAGCTTAAAGCAAAATTAGACAGGCTGAGAAAAAGCTACTGAAGCCAAA 711  
 QY 982 cctgcagatggtctgatgtttcccaatccaggaactacagttggaggctccaagcaaa 1041

Db 712 CAAAAAGTGATGCTGAAGAAGTCGCTCTCAAGCTAAATCGCTGAATTGAAAAATCAA 771  
 QY 1042 ggaagtagtattg 1054  
 Db 772 GTTCATAGACTAG 784  
 RESULT 10  
 US-08-214-164-1  
 Sequence 1, Application US/08214164  
 Patent No. 5728387  
 GENERAL INFORMATION:  
 APPLICANT: BRILES, DAVID E.  
 APPLICANT: YOTTER, JANET L.  
 TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN  
 NUMBER OF SEQUENCES: 3  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Mattare, Ltd  
 STREET: Suite 1203, 2001 Jefferson Davis Highway  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/214,164  
 FILING DATE: 17-MAR-1994  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/656,773  
 FILING DATE: 15-FEB-1991  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Berkstresser, Jerry W.  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: 6102-137  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 521-0813  
 TELEX: LURPAT WASHINGTON  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2085 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 1..1983  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 127..1983  
 US-08-214-164-1

Query Match 3.4%; Score 50.6; DB 1; Length 2085;  
 Best Local Similarity 49.8%; Pred. No. 4.3e-05;  
 Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;  
 QY 745 atagagaagaatggaatgcgattgggagatgcatatttgcaggacagaaagcgtagtga 804  
 DB 472 ACAGACAAGCGCGAAGAACGACGACGACGACGACGACGACGACGACGACGACGACG 531  
 QY 805 gctgtagaaaatgctaaatcctaataacagataaagcagataaagcagataaagcagca 864  
 DB 532 GAAGAGGCAAAACTAAATTTAACTGTTTCGAGCAATGGTAGTTCCTGAGCCAGCAG 591  
 QY 865 atcgtactgctaagacacaaaatagctgaagctcagaaaaagttcccca 921







FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOWALSKI, Thomas J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454312-2064  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2085 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
STRAIN: Streptococcus pneumoniae Rxi  
IMMEDIATE SOURCE:  
CLONE: JY2008  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1..2085  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(127..1983, 1987..1992)  
LOCATION: ..2025, 2029..2031, 2035..  
US-08-467-852A-1

[illegible]

	QY	922	atcttcaagaagcggaacaataattggttaatatcaggcttgagaagatctttaaaaatatcaa	981
	Dd	652	AAACTAGAAGCAAGCTAAGCCAAAATTGAGAGAGCTCAGAAAAAAGCTACTGAAGCCCAA	711
	QY	982	cctgcgatagtgttgtgatgttcacaaatccaaggaaactcacagttggaggctccaaagcaaac	1041
	Dd	712	C AAAAAGTGGATGTTCGAAGAAGTCGCCTCCAAGCTAAAAATCGCTGAATTGGAAAAATCAA	771
	QY	1042	ggaagtagtatgtt 1054	
	Dd	772	GTTCATAGACTAG 784	
 RESULT 14 US-08-468-718-1 ; Sequence 1, Application US/08468718 ; Patent No. 5871943 GENERAL INFORMATION: ; APPLICANT: Briles, David E ; APPLICANT: Yother, Janet L ; APPLICANT: McDaniel, Larry S TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL TITLE OF INVENTION: PROTEIN NUMBER OF SEQUENCES: 6 CORRESPONDENCE ADDRESS:				

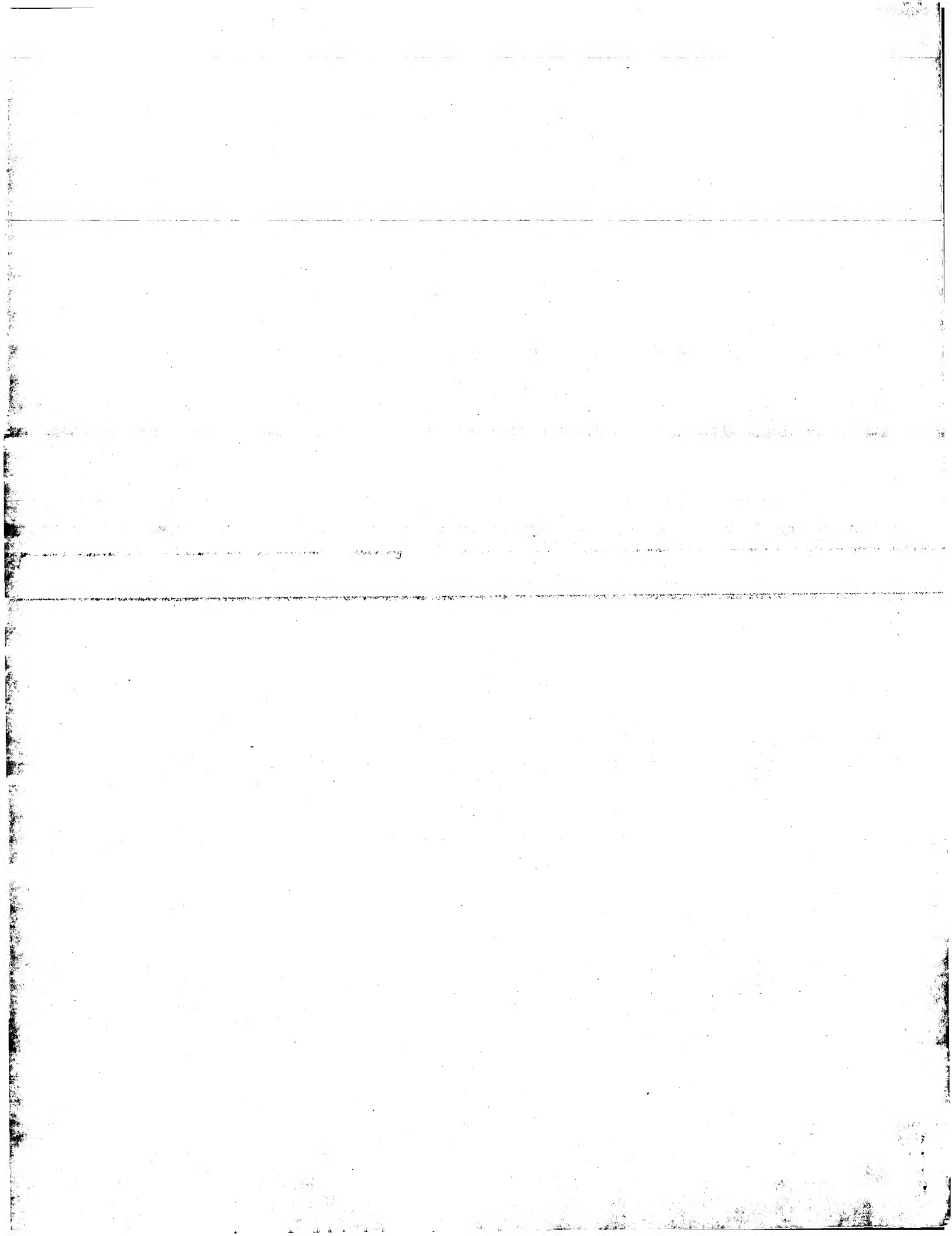


۱۰۰



Search completed: February 7, 2002, 15:53:51  
Job time: 345 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: February 7, 2002, 21:32:15 ; Search time 7419.31 Seconds  
(without alignments)  
2129.079 Million cell updates/sec

Title: US-09-391-606-3

Perfect score: 1470

Sequence: 1 atgtgtaactctattgtgcc.....accatcatcatcatcat 1470

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_hic:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_hic:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rpd:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	45	3.1	861	13	AZ682538
C 2	45	3.1	906	13	AZ550193
C 3	44.4	3.0	484	10	AA314486
C 4	42.8	2.9	856	13	BH132709
C 5	42.6	2.9	460	10	AA672611
C 6	42	2.9	624	11	BF865301
C 7	41.8	2.8	586	13	AZ632588
C 8	41.8	2.8	850	10	AV406016
C 9	41.6	2.8	762	10	BE661568
C 10	41.4	2.8	842	13	AZ536150
C 11	41.4	2.8	845	13	AZ685553
C 12	41.4	2.8	876	13	AZ691916

13	41.2	2.8	616	13	AZ525704
C 14	41.2	2.8	737	10	AU075947
C 15	41.2	2.8	1000	11	BG291509
C 16	40.8	2.8	841	13	AZ676761
C 17	40.6	2.8	450	11	BG893782
C 18	40.6	2.8	527	13	AQ227560
C 19	40.6	2.8	1042	13	CNS0148K
C 20	40.4	2.7	1101	13	CNS00L4E
C 21	39.8	2.7	915	13	CNS07CXR
C 22	39.8	2.7	938	13	CNS006TJ
C 23	39.6	2.7	422	11	RI5357
C 24	39.6	2.7	502	13	B60448
C 25	39.6	2.7	720	10	BE783140
C 26	39.6	2.7	946	11	BG339626
C 27	39.4	2.7	262	10	BB454826
C 28	39.4	2.7	618	11	BI394077
C 29	39.4	2.7	856	13	AZ667395
C 30	39.2	2.7	298	10	AI218089
C 31	39.2	2.7	384	11	RO3975
C 32	39.2	2.7	695	10	BE217675
C 33	39.2	2.7	855	13	AZ688493
C 34	39.2	2.7	863	13	AZ548883
C 35	39.2	2.7	885	13	AZ693152
C 36	39.2	2.7	938	13	AZ549589
C 37	39	2.7	468	13	AZ522521
C 38	39	2.7	576	11	BG602389
C 39	39	2.7	748	11	BG124392
C 40	39	2.7	884	13	AZ136778
C 41	39	2.7	904	13	AZ692954
C 42	39	2.7	935	13	CNS033D4
C 43	39	2.7	1013	10	AL564621
C 44	38.8	2.6	450	13	FR0025683
C 45	38.8	2.6	738	10	AV405991

#### ALIGNMENTS

RESULT 1	AZ682538	861 bp	DNA	GSS	14-DEC-2000
LOCUS	ENT1ALTFF	Entamoeba histolytica	Sheared DNA	Entamoeba histolytica	
DEFINITION	genomic, DNA sequence.				
ACCESSION	AZ682538				
VERSION	AZ682538.1	GI:11819684			
KEYWORDS	GSS.				
SOURCE	Entamoeba histolytica.				
ORGANISM	Eukaryota; Entamoebidae; Entamoeba.				
REFERENCE	1 (bases 1 to 861)				
AUTHORS	Loftus, B., Van Aken, S. and Fraser, C.				
TITLE	Determination of clone end sequences from Entamoeba histolytica				
JOURNAL	HMI:IMSS sheared DNA library				
COMMENT	Unpublished (2000)				
	Contact: Brendan J Loftus				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: entae@igrr.org				
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared				
	DNA library				
	Seq primer: M13-Forward				
	Class: shotgun				
	High quality sequence start: 36				
	High quality sequence stop: 813.				
	Location/Qualifiers				
	1. .861				
	/organism="Entamoeba histolytica"				
	/strain="HMI:IMSS"				
	/db_xref="taxon:5759"				
	/clone_lib="Entamoeba histolytica Sheared DNA"				

FEATURES  
source







Fax: 301 838 0208

REFERENCE 1 (bases 1

Eukaryota; Metazoa: Chordata: Craniata: Vertebrata; Euteleostomi; Mammalia: Eutheria: Rodentia: Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 460)

REFERENCE

AUTHORS

Marra, M., Hillier, J., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,







Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 586)  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0487 row: L column: 17  
Seq primer: CGTTGTAAGCAGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 586.  
FEATURES  
source Location/Qualifiers  
1..586  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUCG1M0487L17"  
/clone\_lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gii4732114|gbiAF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 152 a 127 c 125 g 182 t  
ORIGIN  
Query Match 2.8%; Score 41.8; DB 13; Length 586;  
Best Local Similarity 51.3%; Pred. No. 2.3;  
Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
QY 1011 aggaactacagtgtgagggtcccaagcaaggaagtagtagtattgtagtattcgtttc 1070  
DB 209 AGAAGGAAGACGACGAGCAGCAGCTAGGCGAGAGAACTGGTGGAACTTGAAGCCA 150  
QY 1071 catgctgttagtgcgtgaaatgagacgcctccatttgcgtggtttcgtca 1130  
DB 149 AAACCCACTAGTGTGCTTCGACAGAGACCCAAATTCGATTGCACATTTTGGGCGCTCTG 90  
QY 1131 gatgattcacatgtttacacgaaatcctgattcctcaagctgccacagagactcgc 1190  
DB 89 GGTACTGGGCATGATGAGGAGCCCTAGAAATTCCTGCTCAAGACACCCCAAGACCATCTC 30  
QY 1191 agcaacaagc 1199  
| | | | |

Db 29 AACACAAGC 21  
RESULT 8  
AV406016  
LOCUS  
DEFINITION AV406016 Bombyx mori wing disk C108 5th-instar day-3 larva Bombyx mori cDNA clone wdv30506 T3, mRNA sequence.  
ACCESSION AV406016  
VERSION AV406016.1 GI:6910104  
KEYWORDS EST.  
SOURCE Bombyx mori  
ORGANISM domestic silkworm.  
REFERENCE 1 (bases 1 to 850)  
AUTHORS Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.  
TITLE Bombyx mori cDNA  
JOURNAL Unpublished (2000)  
COMMENT Contact: Mita K  
Genome Research Group  
National Institute of Radiological Sciences  
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan  
Email: kmita@nirs.go.jp  
method: uni-directional, sequence direction: sequenced from T3 primer (5' -> 3')  
Project: 'Silkworm Genome Program in MAFF, and Research for the Future Program in JSPS', see 'SilkBase',  
Future Program in JSPS, see 'SilkBase',  
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.  
FEATURES  
source Location/Qualifiers  
1..850  
/organism="Bombyx mori"  
/strain="C108"  
/db\_xref="taxon:7091"  
/clone="wdv30506"  
/clone\_lib="Bombyx mori wing disk C108 5th-instar day-3 larva"  
/sex="female/male mixed"  
/tissue type="wing disk"  
/dev stage="5th-instar day-3 larva"  
BASE COUNT 322 a 149 c 189 g 190 t  
ORIGIN  
Query Match 2.8%; Score 41.8; DB 10; Length 850;  
Best Local Similarity 47.2%; Pred. No. 2.6;  
Matches 127; Conservative 0; Mismatches 142; Indels 0; Gaps 0;  
QY 697 gcaattgctcaattctttagttgatcagacagatgctacagcgacacagatagagaagat 756  
DB 459 GAAATAGCTGTAAGAAACCTCCCAAGAAACAGAACTACACGCGAAAAAGCAGAAAAATGAT 518  
QY 757 ggaataatgcattggggatgcataatttgcagacagaaacgctagtggaacttagaaaaat 816  
DB 519 GAAGCAGACACCATTTGAACATCTGAAAGTAAGTAAGAGACTAGTCTCTTTAAGAAAAAC 578  
QY 817 gctaattctaataacagcataagcaacatagattcagctaaagcagcaatcgcactcgt 876  
DB 579 TCACCAATTAAAGAAATTAAGAAACTACAAAAAGAGCTACACCTGTTCTCTAGATAAT 638  
QY 877 aagacacaaaatagctgaagctcagaaaaagttcccgactcctccatttccaagaagcg 936  
DB 639 GAAGTGCAGGAAGATGTTGCTAAACAAAGGTTCCACCCGAGAAAGAGCCCTCACCAGATCTA 698  
QY 937 gacaaaatggttaatacagcgtcgagaaaga 965  
DB 699 CAAATTTGGTGGAGATGAACCAAGA 727  
| | | | |

RESULT 9  
BE661568  
LOCUS  
BE661568 762 bp mRNA EST 06-SEP-2000







[illegible]



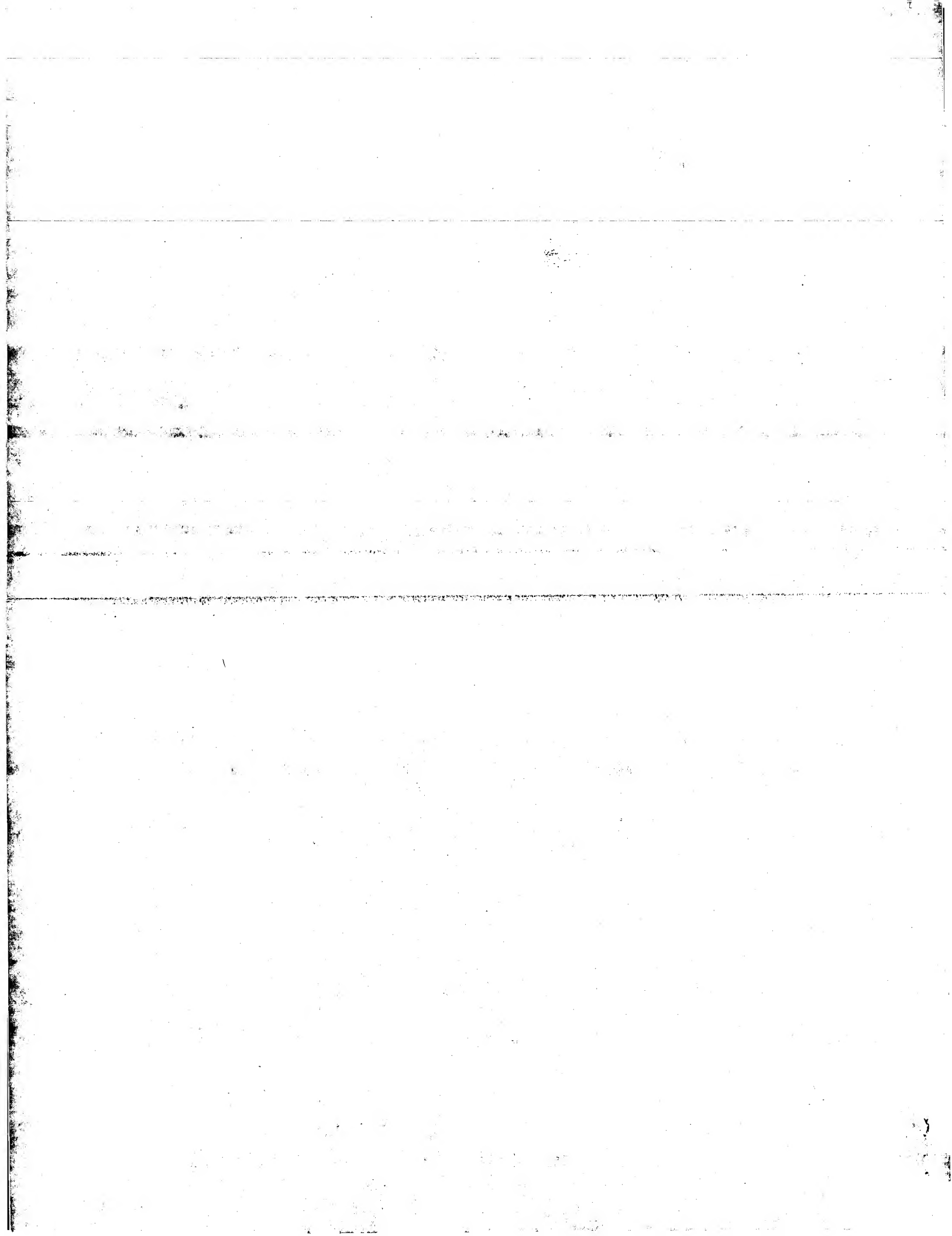
Qy	1000	gttccaaatccaggaactacagtggtgaggtcccaagcaacaaaggaagtagtagtggtagt	1059
Db	500	ATTATAGTCAAGTAACAATTGTTGGATGTTTAAATGGACAATTACATGATTTATTAAAT	559
Qy	1060	att 1062	
Db	560	ATT 562	
RESULT 13			
LOCUS	AZ525704	616 bp DNA GSS	07-MAY-2001
DEFINITION	247Pbc05 Pb MBN #21 Plasmodium berghei genomic 3', DNA sequence.		
ACCESSION	AZ525704		
VERSION	AZ525704.1	GI:13966815	
KEYWORDS	GSS.		
SOURCE	Plasmodium berghei.		
ORGANISM	Plasmodium berghei.		
REFERENCE	1 (bases 1 to 616)		
AUTHORS	Carlton, J.M.-R. and Dame, J.B.		
TITLE	The Plasmodium vivax and P. berghei gene sequence tag projects		
JOURNAL	Parasitol. Today 16 (10), 409 (2000)		
COMMENT	Contact: Dame JB Dept. of Pathobiology, College of Veterinary Medicine University of Florida 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA Tel: 352 392 4700 Fax: 352 392 9704 Email: damej@mail.vetmed.ufl.edu Seq primer: M13(-20) forward Class: Shotgun.		
FEATURES	source	Location/Qualifiers	
	1..616	/organism="Plasmodium berghei"	
		/strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"	
		/db_xref="taxon:5821"	
		/clone_lib="Pb MBN #21"	
		/dev_stage="asexual blood forms"	
		/lab_host="Mus musculus"	
		/note="Vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site_1: EcoRV; Site_2: EcoRV; Genomic DNA was prepared from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei grown in laboratory Swiss white mice. The DNA was purified from contaminating host DNA by Hoechst Dye 33258-Cscl ultracentrifugation and precipitated. Purified DNA was digested with mung bean nuclease in the presence of 36-38% formamide at 50 C, as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the digestion fragments were polished using T4 DNA polymerase, and the fragments size selected in the range 500-2000 bp. These were ligated into the EcoRV-cleaved and dephosphorylated pBluescript SK(+) vector. Recombinant plasmids were used to transform E. coli XL10-Gold host cells."	
BASE COUNT	271 a	88 c	85 g
ORIGIN	172 t		
Query Match 2.8%; Score 41.2; DB 13; Length 616;			
Best Local Similarity 53.8%; Pred. No. 3.4;			
Matches 85; Conservative 0; Mismatches 73; Indels 0; Gaps 0;			
Qy	714	agttgacacagatgctacagcgacagatagagaagatggaaatgcgattgggga	773
Db	45	AGGTGCTGATACATCTCTCAAAATAACCAACACAAACAGATGGACCAATATTATAG	104
Qy	774	tgcataatttcgagcagacagcgtagtgagctgtagaaatgctaaatcctaataacag	833
Db	105	ATCATTATTAGCATTAAGAGAGTGTATTCGAGCTATGCTGATTCAGATAAAATCATATACC	164
us-09-391-606-3.rst			
RESULT 14			
LOCUS	AU075947/c	737 bp mRNA	15-JUL-1999
DEFINITION	AU075947 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium discoideum CDNA clone SSA185, mRNA sequence.		
ACCESSION	AU075947		
VERSION	AU075947.1	GI:5498843	
KEYWORDS	EST.		
SOURCE	Dictyostelium discoideum.		
ORGANISM	Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.		
REFERENCE	1 (bases 1 to 737)		
AUTHORS	Urushihara, H.		
TITLE	Developmental cDNA in Dictyostelium discoideum (1999)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Hideko Urushihara Institute of Biological Sciences University of Tsukuba 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan Email: d402huesakura.cc.tsukuba.ac.jp PROJECT = Dictyostelium discoideum CDNA project in Japan.		
FEATURES	source	Location/Qualifiers	
	1..737	/organism="Dictyostelium discoideum"	
		/strain="AX4"	
		/db_xref="taxon:44689"	
		/clone="SSA185"	
		/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"	
		/dev_stage="slug"	
BASE COUNT	253 a	109 c	87 g
ORIGIN	288 t		
Query Match 2.8%; Score 41.2; DB 10; Length 737;			
Best Local Similarity 51.1%; Pred. No. 3.6;			
Matches 97; Conservative 0; Mismatches 93; Indels 0; Gaps 0;			
Qy	818	ctaaatctaaataacagtagtaagcaacatagattcagctaaagcagcaatcgctactgcta	877
Db	316	CCAAATGAATCAACAATCTATGTAATAATATATCAATTAATGACCACATAATAACCTA	257
Qy	878	agacacaaatgctgaagctcagaaaaagttcccgactctccaattctccaagaagcgg	937
Db	256	TAAAAAATATAACTGTTCTATATAAACTGATGATTCTCTTGAATTTATGAATAAATAT	197
Qy	938	aacaaatggtataacagcgtgagaaagatcttataaaatatcaaacctgcagatggtctg	997
Db	196	AAATGCTGTAATAATTTGTTATCCATAAAATGAAAATATACCATGCGCAACGTGTGA	137
Qy	998	atgttccaaa 1007	
Db	136	ATAATTCCTAA 127	
RESULT 15			
LOCUS	BG291509	1000 bp mRNA	21-FEB-2001
DEFINITION	602385708f1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4514905 5', mRNA sequence.		
ACCESSION	BG291509		
VERSION	BG291509.1	GI:13049457	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 1000)		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		



http://image.llnl.gov  
Plate: LLAM10403 row: 0 column: 02  
High quality sequence start: 12  
High quality sequence stop: 594.

Search completed: February 7, 2002, 21:32:23  
Job time: 20657 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:59:45 ; Search time 685.35 Seconds  
(without alignments)  
1737.543 Million cell updates/sec

Title: US-09-391-606-4

Perfect score: 1389

Sequence: 1 atggttaatcctattgtcc.....aagttctatggtccgagct 1389

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1379.4	99.3	2238	21	AA02066 C. pneumoniae 76 k
2	1359.4	97.9	2156	21	AA02063 Chlamydia pneumoniae
3	1348.4	97.1	1230025	20	AA02065 Nucleotide sequence
4	1347	97.0	1456	21	AA02065 3'-truncated Chlam
5	1155.4	83.2	1852	21	AA02064 5'-truncated Chlam
6	1146.8	10.6	150	19	AA02064 Part of the gene e
7	83	6.0	936	22	AA02063 Oligonucleotide D1
8	83	6.0	936	22	AA02063 Oligonucleotide D1
9	83	6.0	936	22	AA02063 Oligonucleotide D1
10	83	6.0	936	22	AA02063 Oligonucleotide D1
11	83	6.0	936	22	AA02063 Oligonucleotide D1

c	12	83	6.0	938	22	AA02063
	13	80.6	5.8	936	22	AA02063
	14	80.6	5.8	936	22	AA02063
	15	80.6	5.8	936	22	AA02063
	16	80.6	5.8	936	22	AA02063
	17	80.6	5.8	936	22	AA02063
	18	80.6	5.8	938	22	AA02063
	19	70	5.0	1550	21	AAA28411
	20	53.6	3.9	1537	21	AA02063
	21	53.6	3.9	1537	22	AA02063
	22	50.6	3.6	896	15	AA02063
	23	50.6	3.6	946	20	AA02063
	24	50.6	3.6	957	18	AA02063
	25	50.6	3.6	1990	18	AA02063
	26	50.6	3.6	2085	15	AA02063
	27	50.6	3.6	2085	17	AA02063
	28	50.6	3.6	2085	17	AA02063
	29	50.6	3.6	2085	19	AA02063
	30	50.6	3.6	2085	19	AA02063
	31	50.6	3.6	2085	20	AA02063
	32	50.6	3.6	2085	20	AA02063
	33	50.6	3.6	2085	20	AA02063
	34	49	3.5	2085	13	AA02063
	35	43.6	3.1	6744	13	AA02063
	36	41.4	3.0	1038602	20	AA02063
	37	40	2.9	1196	18	AA02063
	38	40	2.9	4590	22	AA02063
	39	39.6	2.9	640	19	AA02063
	40	39.6	2.9	1237	22	AA02063
	41	39.6	2.9	1352	21	AA02063
	42	39.6	2.9	3704	22	AA02063
	43	39.6	2.9	5176	22	AA02063
	44	39.6	2.9	6314	22	AA02063
	45	39.4	2.8	3337	17	AA02063

## ALIGNMENTS

RESULT 1  
AA02066  
ID AA02066 standard; DNA; 2238 BP.

AC AA02066;

DT 26-MAR-2001 (first entry)

DE C. pneumoniae 76 kDa protein truncation mutant fusion gene.

XX 76 kDa protein; bactericidal; diagnosis; prevention;  
KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;  
KW acute respiratory disease; cough; sore throat; hoarseness; fever;  
KW vaccine; immunisation; treatment; truncation mutant; fusion gene;  
KW ds.

OS Chlamydia pneumoniae.

OS Synthetic.

FH Key Location/Qualifiers

FT misc\_feature

FT 1..665

FT /\*tag= a

FT /note= "This part of the sequence is unrelated to

FT C. pneumoniae 76 kDa gene"

FT 766..2238

FT /\*tag= b

FT /product= "Truncated Chlamydia pneumoniae

FT 76KDa protein"

FT 2122..2238

FT /\*tag= c

FT /note= "This part of the sequence is unrelated to

FT C. pneumoniae 76 kDa gene"

PN W0200066739-A2.



Fri Feb 8 08:45:16 2002

```
XX PD 09-NOV-2000.
XX PF 03-MAY-2000; 2000WO-CA00511.
XX PR 03-MAY-1999; 99US-0132270.
XX PR 30-JUN-1999; 99US-0141276.
XX PA (AVET ) AVENTIS PASTEUR LTD.
XX PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX DR WPI: 2000-687542/67.
XX DR P-PSDB; AAY71957.
XX PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX PT useful for vaccinating against Chlamydia infections -
XX PS Claim 32; Fig 3; 112pp; English.
XX CC The present sequence is a DNA coding for a fusion protein comprising a
XX CC truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded
XX CC residues. C. pneumoniae 76 kDa protein is used in the
XX CC diagnosis, prevention and treatment of C. pneumoniae
XX CC infections (e.g. pneumonia, upper respiratory
XX CC tract disease, bronchitis, sinusitis and acute respiratory
XX CC disease such as cough, sore throat, hoarseness, fever; and
XX CC abnormal chest sounds on auscultation). C. pneumoniae sequence
XX CC is also used as vaccines for immunising humans against diseases
XX CC caused by C. pneumoniae.
XX SQ Sequence 2238 BP; 689 A; 459 C; 497 G; 593 T; 0 other;

Query Match 99.3%; Score 1379.4; DB 21; Length 2238;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 193; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atgttgaatcctattggtccaggctctatagacgaacaaagacacaccccccagat 60
DB 766 atgttgaatcctattggtccaggctctatagacgaacaaagacacaccccccagat 825
QY 61 ctcttcgctcaagattgagcgagtcgacgaataaagtcggaagctcaagaata 120
DB 826 ctcttcgctcaagattgagcgagtcgacgaataaagtcggaagctcaagaata 885
QY 121 gcagggtgcggaagctaaagctaaagctaaagctaaagctaaagctaaagctaa 180
DB 886 gcagggtgcggaagctaaagctaaagctaaagctaaagctaaagctaaagctaa 945
QY 181 ttgcgttctcagtgatgaatctcatgagctctggcagataagctgggtattgtctt 240
DB 946 ttgcgttctcagtgatgaatctctcagctctggcagataagctgggtattgtctt 1005
QY 241 aacagctcgtctctactagcagatctgcagacgtggactaacgacagcagccact 300
DB 1006 aacagctcgtctctactagcagatctgcagacgtggactaacgacagcagccact 1065
QY 301 acgctcctccaccacgctgattgattataagactcaagcgaacacagcttacgatact 360
DB 1066 acgctcctccaccacgcttggattataagactcaagcgaacacagcttacgatact 1125
QY 361 atctttaccctcaacatcactagctgacatacagctgctttgttgagcctccaggatgct 420
DB 1126 atctttaccctcaacatcactagctgacatacagctgctttgttgagcctccaggatgct 1185
QY 421 gtcactaataaaggatacagcgctactgatgaggaacacccgaatcgctgcggagtgg 480
DB 1186 gtcactaataaaggatacagcgctactgatgaggaacacccgaatcgctgcggagtgg 1245
QY 481 gaaactaagaatgcgatgaatctaaagtggcgcgcaaatattacagaaattagcgaaatat 540
DB 1246 gaaactaagaatgcgatgaatctaaagtggcgcgcaaatattacagaaattagcgaaatat 1305
```

```
QY 541 gcttcggataaaccagcgattcttgaactctttaggtaaactgaactctctcgaccttta 600
DB 1306 gcttcggataaaccagcgattcttgaactctttaggtaaactgaactctctcgaccttta 1365
QY 601 cagactgctcttctcaatctgtagcaacaataaacaagcagctgagctctttaaagag 660
DB 1366 caggctgctcttctcaatctgtagcaacaataaacaagcagctgagctctttaaagag 1425
QY 661 atgcaagataaaccagtagtcccgagggaacacgctgcgaattgctcaattctttagttgat 720
DB 1426 atgcaagataaaccagtagtcccgagggaacacgctgcgaattgctcaattctttagttgat 1485
QY 721 cagacagatgctacagcgacacagatagaaaagatggaatgcgattgggagatcatat 780
DB 1486 cagacagatgctacagcgacacagatagaaaagatggaatgcgattgggagatcatat 1545
QY 781 ttgcaggacagaaacgctagtgagctgtagaaaaatgctaaatcttaataaacagataaagc 840
DB 1546 ttgcaggacagaaacgctagtgagctgtagaaaaatgctaaatcttaataaacagataaagc 1605
QY 841 aacatagattcagctaaagcagcaatcgctactgctaaagacacacaaatagctgaagctcag 900
DB 1606 aacatagattcagctaaagcagcaatcgctactgctaaagacacacaaatagctgaagctcag 1665
QY 901 aaaaagtcccgactctccaattcttcaagaagcggaaacaaatggttaatacaggtgag 960
DB 1666 aaaaagtcccgactctccaattcttcaagaagcggaaacaaatggttaatacaggtgag 1725
QY 961 aaagatcttaaaaaatacacaacctgcagatggtttctgattgttccaaatccagaactaca 1020
DB 1726 aaagatcttaaaaaatacacaacctgcagatggtttctgattgttccaaatccagaactaca 1785
QY 1021 gttgagagctccaaagcaacaggaagtagtattggttagtattctgtttccatgctgtta 1080
DB 1786 gttgagagctccaaagcaacaggaagtagtattggttagtattctgtttccatgctgtta 1845
QY 1081 gatgatgctgaataatgagacgcttccattttgatgtctgtgtgtgtgtgtgtgtgtgtgtgt 1140
DB 1846 gatgatgctgaataatgagacgcttccattttgatgtctgtgtgtgtgtgtgtgtgtgtgt 1905
QY 1141 atgttcaatacggaaaatccctgattctcaagctgcccaacagagagctcgcagacaaagct 1200
DB 1906 atgttcaatacggaaaatccctgattctcaagctgcccaacagagagctcgcagacaaagct 1965
QY 1201 agagcagcgaaacgctggagatgacagtgctgctgcagcgctggcagatgctgcagaaa 1260
DB 1966 agagcagcgaaacgctggagatgacagtgctgctgcagcgctggcagatgctgcagaaa 2025
QY 1261 gctttagaagcggctctaggttaaagctggcacaacacagggcactactcaatgtcttgaga 1320
DB 2026 gctttagaagcggctctaggttaaagctggcacaacacagggcactactcaatgtcttgaga 2085
QY 1321 cagatcgctctgctgctgtgtgtgtgagcgcagagagctcccccgcgtgcagaaagttctatgg 1380
DB 2086 cagatcgctctgctgctgtgtgtgtgagcgcagagagctcccccgcgtgcagaaagttctatgg 2145
QY 1381 atcccgagct 1389
DB 2146 atcccgagct 2154
```

RESULT 2

```
AAD02063
ID AAD02063 standard; DNA; 2156 BP.
XX
AC AAD02063;
XX
DT 26-MAR-2001 (first entry)
XX
DE Chlamydia pneumoniae 76 kDa full-length protein encoding DNA.
KW 76 kDa protein; bactericidal; diagnosis; prevention; treatment;
```



KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;  
 KW acute respiratory disease; cough; sore throat; hoarseness; fever;  
 KW vaccine; immunisation; ds.  
 XX Chlamydia pneumoniae.

XX Key Location/Qualifiers  
 FH 101..2056  
 FT /\*tag= a  
 FT /product= "Chlamydia pneumoniae 76kDa protein"  
 XX

PN WO200066739-A2.

XX 09-NOV-2000.

XX 03-MAY-2000; 2000WO-CA00511.

XX 03-MAY-1999; 99US-0132270.

XX 30-JUN-1999; 99US-0141276.

XX (AVET ) AVENTIS PASTEUR LTD.

PI Murdin AD, Oomen RP, Wang J, Dunn P;

XX WPI; 2000-687542/67.

XX P-PSDB; AAY71954.

XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,  
 XX useful for vaccinating against Chlamydia infections -  
 XX

XX Claim 2a; Fig 1; 112pp: English.

XX The present sequence is a DNA coding for Chlamydia pneumoniae  
 CC 76 kDa protein. C. pneumoniae 76 kDa protein  
 CC is used in the diagnosis, prevention and treatment  
 CC of C. pneumoniae infections (e.g. pneumonia, upper respiratory  
 CC tract disease, bronchitis, sinusitis and acute respiratory  
 CC disease such as cough, sore throat, hoarseness, fever; and  
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence  
 CC is also used as vaccines for immunising humans against diseases  
 CC caused by C. pneumoniae.  
 XX

XX Sequence 2156 BP; 672 A; 461 C; 471 G; 552 T; 0 other;

Query Match		97.9%;	Score 1359.4;	DB 21;	Length 2156;
Best Local Similarity		99.5%;	Pred. No. 0;		
Matches 1374;		Conservative	0;	Mismatches	6;
				Indels	1;
				Gaps	1;
Qy	1	atggttaactctattggtccaggtcctatagacgaaacagacgacacacccctccgcagat	60		
Db	101	atggttaactctattggtccaggtcctatagacgaaacagacgacacacccctccgcagat	160		
Qy	61	ctttctgctcaagattgagcgagtgacgaaataagtgcggaagctcaagaata	120		
Db	161	ctttctgctcaagattgagcgagtgacgaaataagtgcggaagctcaagaata	220		
Qy	121	gcaggtgcggaagctaaagaaatctaaagacgattctgtagacgagtgagcattc	180		
Db	221	gcaggtgcggaagctaaagaaatctaaagacgattctgtagacgagtgagcattc	280		
Qy	181	ttggtttctgagtgaaatgctctcatagtgctggcagataagctgggtattgctctagt	240		
Db	281	ttggtttctgagtgaaatgctctcatagtgctggcagataagctgggtattgctctagt	340		
Qy	241	aacagctgcttcttactagcagatctgacgctggactcaacgacgacgacgacacct	300		
Db	341	aacagctgcttcttactagcagatctgacgctggactcaacgacgacgacgacacct	400		
Qy	301	acgctctctccaccacgctgtgattatagactcaagcgcaaacagcttactgatact	360		
Db	401	acgctctctccaccacgctgtgattatagactcaagcgcaaacagcttactgatact	460		

Qy	361	atctttactcaacatcactagctacatacacaggtgctttgtgagctccagatgct	420
Db	461	atctttactcaacatcactagctacatacacaggtgctttgtgagctccagatgct	520
Qy	421	gtcactataaaggatacacagcgctactgataggaaacgcgaatcgctcgagtg	480
Db	521	gtcactataaaggatacacagcgctactgataggaaacgcgaatcgctcgagtg	580
Qy	481	gaaactaagaatccgagtgcaattaaagtggcgcgcaaatcacagaattacgaatat	540
Db	581	gaaactaagaatccgagtgcaattaaagtggcgcgcaaatcacagaattacgaatat	640
Qy	541	gcttcggataacaaagcattcttgaactctttagtgtaactgactctcttcgacctta	600
Db	641	gcttcggataacaaagcattcttgaactctttagtgtaactgactctcttcgacctta	700
Qy	601	cagactgctcttctccaatctgtgacaaataacaaagcgctgagctctttaaagag	660
Db	701	cagactgctcttctccaatctgtgacaaataacaaagcgctgagctctttaaagag	760
Qy	661	atgcaagataacccagtagtcccgaggaacgcctgcaattgctcaattcttagttgat	720
Db	761	atgcaagataacccagtagtcccgaggaacgcctgcaattgctcaattcttagttgat	820
Qy	721	cagacagatgctacagcgacacagatagagaaagatgaaatgcgattgggagtcata	780
Db	821	cagacagatgctacagcgacacagatagagaaagatgaaatgcgattgggagtcata	880
Qy	781	tttgaggacagaaacgctagtgagctgtagaaatgctaaatcctaataacagataaagc	840
Db	881	tttgaggacagaaacgctagtgagctgtagaaatgctaaatcctaataacagataaagc	940
Qy	841	aactagattcagctaaagcagcaaatcgctactgctaaagacacaaatagctgaagctcag	900
Db	941	aactagattcagctaaagcagcaaatcgctactgctaaagacacaaatagctgaagctcag	1000
Qy	901	aaaaagttcccgactctccaatcttcaagaagcggaacaaatggttaacagctgag	960
Db	1001	aaaaagttcccgactctccaatcttcaagaagcggaacaaatggttaacagctgag	1060
Qy	961	aaagattcttaaaaaatacaaacctgcagatggttctgattgttccaaatccaggaactaca	1020
Db	1061	aaagattcttaaaaaatacaaacctgcagatggttctgattgttccaaatccaggaactaca	1120
Qy	1021	gttgaggctcccaagcaacaaaggaagtagtattgtagtattggtttccatgctgta	1080
Db	1121	gttgaggctcccaagcaacaaaggaagtagtattgtagtattggtttccatgctgta	1180
Qy	1081	gatgatgctgaaaaatgagacgcttccattttgattgctgggtttcgtcagatgattcac	1140
Db	1181	gatgatgctgaaaaatgagacgcttccattttgattgctgggtttcgtcagatgattcac	1240
Qy	1141	atgttcaatacggaaaaatcctgattctcaagctgcccaacagagctcgacacacagct	1200
Db	1241	atgttcaatacggaaaaatcctgattctcaagctgcccaacagagctcgacacacagct	1300
Qy	1201	agagcagcgaaagcgctggagatgacagtgctgctgacgctggcagctgacagataa	1260
Db	1301	agagcagcgaaagcgctggagatgacagtgctgctgacgctggcagataa	1360
Qy	1261	gctttagaaagcgctctaggttaaaagctgggcaacaaacaggggcataactcaatcttggga	1320
Db	1361	gctttagaaagcgctctaggttaaaagctgggcaacaaacaggggcataactcaatcttggga	1420
Qy	1321	caagctgcttctgctgctggtgtgagcgagag-tcctcccgctgcagcagattctatg	1379
Db	1421	caagctgcttctgctgctggtgtgagcgagag-tcctcccgctgcagcagattctatg	1480
Qy	1380	g 1380	
Db	1481	g 1481	



RESULT 3  
 AAX91990/c  
 ID AAX91990 standard; DNA; 1230025 BP.  
 XX AC AAX91990;  
 XX DT 13-SEP-1999 (first entry)  
 XX DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
 XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 XX KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 XX KW vaccine; neutralising epitope; ss.  
 XX OS Chlamydia pneumoniae.  
 XX PN W09927105-A2.  
 XX PD 03-JUN-1999.  
 XX PF 20-NOV-1998; 98WO-IB01890.  
 XX PR 04-NOV-1998; 98US-0107078.  
 XX PR 21-NOV-1997; 97FR-0014673.  
 XX PA (GEST ) GENSET.  
 XX PI Griffais R;  
 XX DR WPI; 1999-357842/30.  
 XX PT Genome sequence of Chlamydia pneumoniae  
 XX PS Claim 1; Page 291-611; 1912pp; English.  
 XX CC The present sequence represents the complete genome of Chlamydia pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.  
 XX SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;  
 Query Match 97.1%; Score 1348.4; DB 20; Length 1230025;  
 Best Local Similarity 99.4%; Pred. No. 0;  
 Matches 1374; Conservative 0; Mismatches 6; Indels 2; Gaps 2;  
 QY 1 atgggttaactctattggtccaggtcctatagacgaagaaacagacacacaccccgagat 60  
 Db 828598 ATGGTTAATCCTATTGTCGAGTCTCTATAGACGAACAGACGACACCTCCCGCAGAT 828539  
 QY 61 ctctctctcaaggattgagcgagtgagcgaataaagagtgaggagctcaagaata 120  
 Db 828538 CTCTCTCTCAAGATTGAGGCGAGTGCAGCAAAATAGAGTGCAGGAGCTCAAGAAATA 828479  
 QY 121 gcaaggcggaagctaaagcctaagataccttaagaccgattctgtagagcgatggagcattc 180  
 Db 828478 GCAGGTGCGGAAGCTAAGACCTAAAGAAATCTTAAGACCGATTCTGTAGAGCGATGGAGCATC 828419  
 QY 181 ttgcgttctcagtgaaatgctcatgagctgcagataagctgggattgcttctagt 240  
 Db 828418 TTGCGTTCTGCAGGAAATGCTCTCATAGTCTGGCAGATTAAGTGGGTATGCTTCTAGT 828359  
 QY 241 aacagctcgtcttctactagcagatctgcagagcgtggagctcaacagcagcgacacct 300  
 Db 828358 AACAGCTCGTCTTCTACTAGCAGATCTGCAGACGTGGACTCAACGACGACGACCGACCT 828299

QY 301 acgcctctccaccacacgctctgatattataagaactcaagcgcaaacagcttacgataact 360  
 Db 828298 ACGCTCTCTCCACCACCGTTTGTATGATTATAAGACTTAAGAGCTTAAGCAGCAACAGCTTACGATACT 828239  
 QY 361 atctttacctcaacatcactagctgacatacagcgctcttggtagcctccagatgct 420  
 Db 828238 ATCTTTACCTCAACATCACTAGCTGACATACAGGCTGCTTTGGTAGCCCTCCAGGATGCT 828179  
 QY 421 gtcactaataaaagatacagcgctactgatgagaaacccgcaatcgcctcgaggtgg 480  
 Db 828178 GTCACATATATAAAGATACAGCGCTACTGATGAGAAACCGCAATCTCGCGGAGTGG 828119  
 QY 481 gaaactaaagaatccgagatgcaattaaagtggcgcaaatatacagaattagcagaatat 540  
 Db 828118 GAAACTAAGAATCCGATGCGAGTTAAAGTTGCGCGCAAAATTACAGAATTACGCAAAATAT 828059  
 QY 541 gcttcgataaaccagcattcttgactcttttaggttaaaactgactccttcgacctetta 600  
 Db 828058 GCTTCGGATTAACCAAGCGATTCTTGACTCTTTAGGTAAACTGACTTCTCTCGACCTCTTA 827999  
 QY 601 cagactgctctctccaaatctgtagcaaaacaataaacaagcagctgagcttcttaagag 660  
 Db 827998 CAGCTGCTCTTCTCCAATCTGTAGCAACAATAAACAAGCAGCTGAGCTTCTTAAGAG 827939  
 QY 661 atcagaataaaccagtagtcccgagggaacacccctgcgaatttctcaattttagttgat 720  
 Db 827938 ATGCAGATAACCCAGTAGTCCAGGGAACCGCTGCAATGCTCAATCTTTAGTTGAT 827879  
 QY 721 cagacagatctacagcgacacagatagagaagatggaatgcgattgggagtcatat 780  
 Db 827878 CAGACAGATGCTACAGCGACACAGATAGAGAAAGATGGAATGCGATTAGGATGCATAT 827819  
 QY 781 ttgcaggacagaacgctagtggagctgtgaaataatgctaaataataacagtatagc 840  
 Db 827818 TTTCAGGACAGAACCGCTAGTGGAGCTGTAGAAATGCTAAATTAATAACAGTATAGC 827759  
 QY 841 aacatagattcagtaaaagcagaacacgctactctactctgaacacacaaatagctgaagctcag 900  
 Db 827758 AACATAGATTCAAGTAAGACGCAATCGTACTCTTAGACACAAATAGCTGAAGCTCAG 827699  
 QY 901 aaaaagttcccccactctccaaattctcaagaagcgcaacaaatggtaatacagcgctgag 960  
 Db 827698 AAAAAGTTCGCCGACTCTCCAATTTCTCAAGAGCGGAACAAATGGTAATACAGGCTGAG 827639  
 QY 961 aaagatcttaaaataatacaacccctgcagatggttctgattgttccaaatccaggaactaca 1020  
 Db 827638 AAAGATCTTAAAAATATCAAACTCGAGATGGTTCTGTATGTTCCAAATCCAGGAACATACA 827579  
 QY 1021 gttgagagctccaagcaaacgaagtagtattggttagtattctgttttccatgctgta 1080  
 Db 827578 GTTGGAGGCTCCAAGCAACAGGAAGTAGTATTGGTAGTATTTCGTGTTTCCATGCTGTTA 827519  
 QY 1081 gatgatgctgaataatgagaccgcttccattttgatgctggttgggtttcgagatgatccac 1140  
 Db 827518 GATGATGCTGAAATGAGACCGCTTCCATTTTGTATGCTGGGTTTCGTCAGATGATTCAC 827459  
 QY 1141 atgttcataacgaaataatcctgattctcaagctgcccaacagagagctcgagcacaagct 1200  
 Db 827458 ATGTTCAATACGGAAATTCCTGATTCTCAAGCTGCCCAACAGGAGCTCGCAGCACAGCT 827399  
 QY 1201 aqagcagcgaagcgcgctggagatgacagtgctgctgcagcgctggcagatgctcagaaa 1260  
 Db 827398 AGAGCAGCGAAAGCCGCTGGAGATGACAGTGTCTGTGACGCTGGCAGATGCTCAGAAA 827339  
 QY 1261 gctttagaagcggtctcagtgaaagctgggcaaacagcgccataactcaatgctcttgga 1320  
 Db 827338 GCTTTAGAAGCGGCTCTAGTAAAGCTGGGCAACAGAGGGCATACTCAATGCTTTAGGA 827279  
 QY 1321 cagatcgctctgctgctgctg-ttgtgagcgagag-tcctcccgctgcagcagagttctat 1378  
 Db 827278 CAGATCGCTTCTGCTGCTGCTTTGTGAGCGCCAGAGTTCTCTCCGCTGCAGCAAGTTCTAT 827219  
 QY 1379 gg 1380



Db 827218 AG 827217

RESULT 4

AAD02065

XX AAD02065 standard; DNA; 1456 BP.

XX AAD02065;

XX 26-MAR-2001 (first entry)

XX 3'-truncated Chlamydia pneumoniae 76 kDa protein encoding DNA.

XX 76 kDa protein; bactericidal; diagnosis; prevention;

XX pneumonia; upper respiratory tract disease; bronchitis; sinusitis;

XX acute respiratory disease; cough; sore throat; hoarseness; fever;

XX vaccine; immunisation; treatment; truncation mutant; ds.

XX Chlamydia pneumoniae.

XX Synthetic.

XX Key Location/Qualifiers

XX CDS 101..1456

XX /\*tag= a

XX /product= "3'-truncated Chlamydia pneumoniae

XX 76kDa protein"

XX /note= "The coding region does not include stop codon"

XX /partial

XX WO200066739-A2.

XX 09-NOV-2000.

XX 03-MAY-2000; 2000WO-CA00511.

XX 03-MAY-1999; 99US-0132270.

XX 30-JUN-1999; 99US-0141276.

XX (AVET ) AVENTIS PASTEUR LTD.

XX Murdin AD, Oomen RP, Wang J, Dunn P;

XX WPI; 2000-687542/67.

XX P-PSDB; AAY71956.

XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,

XX useful for vaccinating against Chlamydia infections -

XX Claim 2c; Page 102-104; 112pp; English.

XX The present sequence is a DNA coding for 3'-truncated

XX Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa

XX protein is used in the

XX diagnosis, prevention and treatment of C. pneumoniae

XX infections (e.g. pneumonia, upper respiratory

XX tract disease, bronchitis, sinusitis and acute respiratory

XX disease such as cough, sore throat, hoarseness, fever; and

XX abnormal chest sounds on auscultation). C. pneumoniae sequence

XX is also used as vaccines for immunising humans against diseases

XX caused by C. pneumoniae.

XX Sequence 1456 BP; 452 A; 308 C; 331 G; 365 T; 0 other;

XX SQ

Query Match 97.08; Score 1347; DB 21; Length 1456;

Best Local Similarity 99.68; Pred. No. 0;

Matches 1350; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 atgggttaatcctattgtccaggctctctatagacgaacacagacacacaccccgagat 60

Db 101 atgggttaatcctattgtccaggctctctatagacgaacacagacacaccccgagat 160

61 ctttctgctcaaggattgagcgagtgacgcaataagatgagcgagctcaagaata 120

Db 161 ctttctgctcaaggattgagcgagtgacgcaataagatgagcgagctcaagaata 220

Qy 121 gcaggtgcggaagctaaagcctaaagaattcaagacgcgattctgttagagcgatggagcattc 180

Db 221 gcaggtgcggaagctaaagcctaaagaattcaagacgcgattctgttagagcgatggagcattc 280

Qy 181 ttggttctgcaagtgaatgctctctatgagctgagcagataaagctgggtattctctctagt 240

Db 281 ttggttctgcaagtgaatgctctctatgagctgagcagataaagctgggtattctctctagt 340

Qy 241 aacagctgcttcttactagcagatctgcagacgtggagcctaacacacagcagcaccact 300

Db 341 aacagctgcttcttactagcagatctgcagacgtggagcctaacacacagcagcaccact 400

Qy 301 acgctctctccacccagctctgattataagacgtcaagcgaacacagcttaccgatact 360

Db 401 acgctctctccacccagcttctgattataagacgtcaagcgaacacagcttaccgatact 460

Qy 361 atctttactcaacatacactagctgacacagctgcttctgtgagcctccagatgct 420

Db 461 atctttactcaacatacactagctgacacagctgcttctgtgagcctccagatgct 520

Qy 421 gtcactataaagatacagcgctctgaggaacacgcgaatcgctcgagtgag 480

Db 521 gtcactataaagatacagcgctctgaggaacacgcgaatcgctcgagtgag 580

Qy 481 gaactaagaatgccgatgcaattaaagtggcgcaaatcacagaatttagcgaatat 540

Db 581 gaactaagaatgccgatgcaattaaagtggcgcaaatcacagaatttagcgaatat 640

Qy 541 gcttcgataaaccagcattcttgaactctttagtaactgaactctctcagcctcta 600

Db 641 gcttcgataaaccagcattcttgaactctttagtaactgaactctctcagcctcta 700

Qy 601 cagactgctcttctccaatctgtagcaacaaatacaaacgagctgagcttcttaagag 660

Db 701 cagactgctcttctccaatctgtagcaacaaatacaaacgagctgagcttcttaagag 760

Qy 661 atgcaagataaccagtagtcccgaggaacacgcgctgaattgctcaatctttagtgat 720

Db 761 atgcaagataaccagtagtcccgaggaacacgcgctgaattgctcaatctttagtgat 820

Qy 721 cagacagatgctacagcagacagatagaaagatgaaatgcgcatgggagtcacat 780

Db 821 cagacagatgctacagcagacagatagaaagatgaaatgcgcatgggagtcacat 880

Qy 781 ttgagagacagacgctagtgagctgtagaaatgctaaatcttaataacagataagc 840

Db 881 ttgagagacagacgctagtgagctgtagaaatgctaaatcttaataacagataagc 940

Qy 841 aacatagattcagctaaagcagcaatcgtctactgttagagacacaaatagctgaagctcag 900

Db 941 aacatagattcagctaaagcagcaatcgtctactgttagagacacaaatagctgaagctcag 1000

Qy 901 aaaaagttcccgactctccaatcttcaagagcggaacaaatggtaatacagctcag 960

Db 1001 aaaaagttcccgactctccaatcttcaagagcggaacaaatggtaatacagctcag 1060

Qy 961 aagatcttaaaaaatcatcaaacctgcagatggtctgtattgttcccaatccaggaactaca 1020

Db 1061 aagatcttaaaaaatcatcaaacctgcagatggtctgtattgttcccaatccaggaactaca 1120

Qy 1021 gttgagagctccaagcaacaaagagtagtattgttagtattcggttccatgctgttta 1080

Db 1121 gttgagagctccaagcaacaaagagtagtattgttagtattcggttccatgctgttta 1180

Qy 1081 gatgatgctgaataatgagacccctccattttgattgctggttcttcgctcagatgattcac 1140

Db 1181 gatgatgctgaataatgagacccctccattttgattgctggttcttcgctcagatgattcac 1240

Qy 1141 atgttcaatacgggaataatcctgattctcaagctgcccacagagagctcgagcagcaagct 1200

Db 827218 AG 827217

RESULT 4

AAD02065

XX AAD02065 standard; DNA; 1456 BP.

XX AAD02065;

XX 26-MAR-2001 (first entry)

XX 3'-truncated Chlamydia pneumoniae 76 kDa protein encoding DNA.

XX 76 kDa protein; bactericidal; diagnosis; prevention;

XX pneumonia; upper respiratory tract disease; bronchitis; sinusitis;

XX acute respiratory disease; cough; sore throat; hoarseness; fever;

XX vaccine; immunisation; treatment; truncation mutant; ds.

XX Chlamydia pneumoniae.

XX Synthetic.

XX Key Location/Qualifiers

XX CDS 101..1456

XX /\*tag= a

XX /product= "3'-truncated Chlamydia pneumoniae

XX 76kDa protein"

XX /note= "The coding region does not include stop codon"

XX /partial

XX WO200066739-A2.

XX 09-NOV-2000.

XX 03-MAY-2000; 2000WO-CA00511.

XX 03-MAY-1999; 99US-0132270.

XX 30-JUN-1999; 99US-0141276.

XX (AVET ) AVENTIS PASTEUR LTD.

XX Murdin AD, Oomen RP, Wang J, Dunn P;

XX WPI; 2000-687542/67.

XX P-PSDB; AAY71956.

XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,

XX useful for vaccinating against Chlamydia infections -

XX Claim 2c; Page 102-104; 112pp; English.

XX The present sequence is a DNA coding for 3'-truncated

XX Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa

XX protein is used in the

XX diagnosis, prevention and treatment of C. pneumoniae

XX infections (e.g. pneumonia, upper respiratory

XX tract disease, bronchitis, sinusitis and acute respiratory

XX disease such as cough, sore throat, hoarseness, fever; and

XX abnormal chest sounds on auscultation). C. pneumoniae sequence

XX is also used as vaccines for immunising humans against diseases

XX caused by C. pneumoniae.

XX Sequence 1456 BP; 452 A; 308 C; 331 G; 365 T; 0 other;

XX SQ

Query Match 97.08; Score 1347; DB 21; Length 1456;

Best Local Similarity 99.68; Pred. No. 0;

Matches 1350; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 atgggttaatcctattgtccaggctctctatagacgaacacagacacaccccgagat 60

Db 101 atgggttaatcctattgtccaggctctctatagacgaacacagacacaccccgagat 160







Db	961	tctcaagctgcccacacaggaagctcgacacaaagctagagcagcgaaagcgcgtgagat	1020
QY	1225	gacagtgtctgtcgacgctggcagatgctcagaagcttttagaagcggtcttaggtaaa	1284
Db	1021	gacagtgtctgtcgacgctggcagatgctcagaagcttttagaagcggtcttaggtaaa	1080
QY	1285	gctgggcaacaacaggggcatactcaatgcttttggacagatcgcttctgctgctgttg	1344
Db	1081	gctgggcaacaacaggggcatactcaatgctttttagacagatcgcttctgctgctgttg	1140
QY	1345	agcgcaggag-tctcccgctgcagcaagttctatgg	1380
Db	1141	agcgcaggagttcttcccgctgcacaaagttctatg	1177

RESULT	6	
AAVi6207		
ID	AAVi6207 standard; DNA; 150 BP.	
XX		
XX	AAVi6207;	
XX		
XX	28-MAY-1998 (first entry)	
DT		
XX		
XX	Part of the gene encoding the 76 kDa cystein rich OMP.	
DE		
XX		
XX	Cysteine rich outer-membrane protein; OMP; 60 kDa OMP; 76 kDa OMP;	
KW	hybridisation; amplification; assay; detection; ds.	
KK		
XX		
XX	Chlamydia pneumoniae.	
OS		
XX		
XX	W09746709-A2.	
PN		
XX		
XX	11-DEC-1997.	
XX		
XX	04-JUN-1997; 97WO-US09673.	
PF		
XX		
XX	06-JUN-1996; 96US-0659473.	
PR		
XX		
XX	(ABEO ) ABBOTT LAB.	
PA		
XX		
XX	Cerney MB;	
PI		
XX		
XX	WPI; 1998-042215/04.	
DR		
XX		
XX	Chlamydia pneumoniae derived oligonucleotides - used as primers or	
PT	probes for specific and sensitive detection	
PT		
XX		
XX	Disclosure; Page 24; 28pp; English.	
PS		

AA	The present sequence represents part of the gene encoding the 76 kDa
CC	cystine rich outer-membrane protein (OMP) of <i>Chlamydia pneumoniae</i> .
CC	Novel oligonucleotides AAV16195-206 can function either as PCR primers
CC	or probes. They are used to detect <i>C. pneumoniae</i> in a sample.
CC	Oligonucleotides V161202-04 are specific for the present sequence.
CC	The oligonucleotides can be used in a hybridisation or amplification
CC	based assay for the detection of <i>C. pneumoniae</i> in a test sample.
XX	
SQ	Sequence 150 BP; 43 A; 34 C; 40 G; 33 T; 0 other;
Query Match	10.6%; Score 146.8; DB 19; Length 150;
Best Local Similarity	98.7%; Pred. No. 1e-32;
Matches 148; Conservative	0; Mismatches 2; Indels 0; Gaps 0;

QY 426 taataaaaggatacacggcctactgatgagaaacccgaatcgcgtcggaatggaaac 485  
|||||  
61 taataaaaggatacacggcctactgatgagaaacccgaatcgcgtcggaatggaaac 120  
Db  
QY 486 taagaatgcgagtcaattaaagtggcgc 515

121 taagaatgccgatgcaggttaaagttggcgc 150

RESULT	7
AAF58252/c	
ID	AAF58252 standard; DNA; 936 BP.
XX	
XX	AAF58252;
XX	
DT	24-APR-2001: (first entry)
XX	
DE	Oligonucleotide D1835.
XX	
XX	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
KW	
OS	Synthetic.
XX	
PN	WO200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	
XX	(CLIN-) CLINICAL MICRO SENSORS INC.
PA	
XX	
XX	Umek RM;
XX	
XX	WPI; 2001-159728/16.
DR	
XX	
PT	Nucleic acids containing electron-transfer group, u
PT	hybridization assays, e.g. for genotyping, allowing
PT	a single surface
XX	
XX	Example 6; Page 127; 159pp; English.
PS	
PS	The present invention relates to a composition comp
XX	acids each containing an electron-transfer group (E
CC	different redox potentials. The invention is used f
CC	detection of nucleic acids, especially of substit
CC	and single-nucleotide polymorphisms, e.g. for genot
CC	monitoring gene expression.
XX	
XX	Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
SQ	



QY 782 ttgcagacagcgtagtgagctgtagaaaatgctaaatctaaatcaacagataaagca 841  
 Db 450 WWWWWW... 391  
 QY 842 acatagattcagctaaagcagcaatcgctactgctaaagacacaaatagctgaagctcaga 901  
 Db 390 WWWWWW... 331  
 QY 902 aaaaagctccagctcctcaatcttcaagaagcgaacaaatggaataacagagctgaga 961  
 Db 330 WWWWWW... 271  
 QY 962 aagatcttaaaatcaaacctcagatggtctgtggttccaaatccagaaactacag 1021  
 Db 270 WWWWWW... 211  
 QY 1022 ttggaggctccaagcaagaagtagtattggttagttcttccatgctgttag 1081  
 Db 210 WWWWWW... 151  
 QY 1082 atgagctgaaaatgagaccgctccatttggatgctggttctcagatgattcaca 1141  
 Db 150 WWWWWW... 91  
 QY 1142 ttttcaatacggaaatcctgattctcaa 1170  
 Db 90 WWWWWW... 62

RESULT 8  
 AAF58254/C  
 ID AAF58254 standard; DNA; 936 BP.  
 XX  
 AC AAF58254;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Oligonucleotide D1875.  
 XX  
 KW Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200107665-A2.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 26-JUL-2000; 2000WO-US20476.  
 XX  
 PR 26-JUL-1999; 99US-0145695.  
 PR 17-MAR-2000; 2000US-0190259.  
 XX  
 PA (CLIN-) CLINICAL MICRO SENSORS INC.  
 XX  
 PI Umek RM;  
 XX  
 DR WPI; 2001-159728/16.  
 XX  
 PT Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface -  
 XX  
 PS Example 6; Page 127; 159pp; English.  
 XX  
 CC The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX

SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;  
 Query Match 6.0%; Score 83; DB 22; Length 936;  
 Best Local Similarity 1.2%; Pred. No. 7.4e-14;  
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;  
 QY 482 aaactaaagaatgcagatgcaattaaagtggcgcgcaaatcacagaatagcgaataatg 541  
 Db 750 WWWWWW... 691  
 QY 542 ctgcgataaacaagcagattcttgaactcttaggttaaactgactcttcacactcttac 601  
 Db 690 WWWWWW... 631  
 QY 602 agactgctcttccaatctgtagcaacaataacaagaagcagctgagcttcttaagaaga 661  
 Db 630 WWWWWW... 571  
 QY 662 tgcaagataaaccagtagtcccgagggaacgcctgcaattgctcaattcttagttgac 721  
 Db 570 WWWWWW... 511  
 QY 722 agacagatgctacagcagacagatagagaagaatggaatgcgattgggagatgcatt 781  
 Db 510 WWWWWW... 451  
 QY 782 ttgcagacagacgcgtagtgagctgtagaaaatgctaaatctaaatcaacagataaagca 841  
 Db 450 WWWWWW... 391  
 QY 842 acatagattcagctaaagcagcaatcgctactgctaaagacacaaatagctgaagctcaga 901  
 Db 390 WWWWWW... 331  
 QY 902 aaaaagctccagctcctcaatcttcaagaagcgaacaaatggaataacagagctgaga 961  
 Db 330 WWWWWW... 271  
 QY 962 aagatcttaaaatcaaacctcagatggtctgtggttccaaatccagaaactacag 1021  
 Db 270 WWWWWW... 211  
 QY 1022 ttggaggctccaagcaagaagtagtattggttagttcttccatgctgttag 1081  
 Db 210 WWWWWW... 151  
 QY 1082 atgagctgaaaatgagaccgctccatttggatgctggttctcagatgattcaca 1141  
 Db 150 WWWWWW... 91  
 QY 1142 ttttcaatacggaaatcctgattctcaa 1170  
 Db 90 WWWWWW... 62  
 RESULT 9  
 AAF58257/C  
 ID AAF58257 standard; DNA; 936 BP.  
 XX  
 AC AAF58257;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Oligonucleotide D1954.  
 XX  
 KW Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200107665-A2.  
 XX



PD 01-FEB-2001.  
 XX 26-JUL-2000; 2000WO-US20476.  
 XX 26-JUL-1999; 99US-0145695.  
 PR 17-MAR-2000; 2000US-0190259.  
 XX (CLIN-) CLINICAL MICRO SENSORS INC.  
 XX UmeK RM;  
 XX WPI; 2001-159728/16.  
 XX Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX Example 6; Page 127; 159pp; English.  
 XX The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 6.0%; Score 83; DB 22; Length 936;  
 Best Local Similarity 1.2%; Pred. NO. 7.4e-14;  
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;  
 QY 482 aaactaagaatccgatgcaattaaagtgtgcgcgcaaatcagaatagcgaatag 541  
 DB 750 WW 691  
 QY 542 ctccggataaccaagcattctgactcttttaggttaaactgactcttcctgaccttac 601  
 DB 690 GWW 631  
 QY 602 agactgctctcctcaatctgtagcaacaataacaaagcagctgactcttctaagaga 661  
 DB 630 WW 571  
 QY 662 tgcaagataaccagtagtcccgaggaaacgcctgcaattgctcaattcttagttgac 721  
 DB 570 WW 511  
 QY 722 agacagatgctacagcagacagatagagaagatggaatgcgattgggagatcatatt 781  
 DB 510 WW 451  
 QY 782 ttgcaggacagacgtagtgagctgtagaaatgctaaatcctaataacagataagca 841  
 DB 450 WW 391  
 QY 842 acatagattcagctaaacagcagcaatcgctactgtaagacacaaatagctgagctcaga 901  
 DB 390 WW 331  
 QY 902 aaaaattcccgactctcattctcagaagcgaacaaatgtaatacaggctgaga 961  
 DB 330 WW 271  
 QY 962 aagattctaaataatcaaacctcgatggtctctgattctcacaatccaggaactacag 1021  
 DB 270 WW 211  
 QY 1022 ttggaggctccaagcaacagaagatgattgtagtattgctgttccatgctgttag 1081  
 DB 210 WW 151

QY 1082 atgatgtgaaatgagaccgcttcattgtatgtctgggtttcgtcagatgattcaca 1141  
 DB 150 WW 91  
 QY 1142 tgttcaatacaggaataatctctgattctcaa 1170  
 DB 90 WW 62  
 RESULT 10  
 AAF58259/C  
 ID AAF58259 standard; DNA; 936 BP.  
 XX AAF58259;  
 XX 24-APR-2001 (first entry)  
 XX Oligonucleotide D2004.  
 XX Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 XX Synthetic.  
 OS WO200107665-A2.  
 PN 01-FEB-2001.  
 PD 26-JUL-2000; 2000WO-US20476.  
 PF 26-JUL-1999; 99US-0145695.  
 PR 17-MAR-2000; 2000US-0190259.  
 XX (CLIN-) CLINICAL MICRO SENSORS INC.  
 PA UmeK RM;  
 XX WPI; 2001-159728/16.  
 DR Nucleic acids containing electron-transfer group, useful as labels in  
 XX hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX Example 6; Page 128; 159pp; English.  
 XX The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 6.0%; Score 83; DB 22; Length 936;  
 Best Local Similarity 1.2%; Pred. NO. 7.4e-14;  
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;  
 QY 482 aaactaagaatccgatgcaattaaagtgtgcgcgcaaatcagaatagcgaatag 541  
 DB 750 WW 691  
 QY 542 ctccggataaccaagcattctgactcttttaggttaaactgactcttcctgaccttac 601  
 DB 690 GWW 631  
 QY 602 agactgctctcctcaatctgtagcaacaataacaaagcagctgactcttctaagaga 661  
 DB 630 WW 571  
 QY 662 tgcaagataaccagtagtcccgaggaaacgcctgcaattgctcaattcttagttgac 721



[illegible]

RESULT	11
AAF58262/C	
ID	AAF58262 standard; DNA; 936 BP.
XX	
AC	AAF58262;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D2007.
XX	
KW	Electron-transfer group; ETW; mismatch; genotyping;
KW	gene expression; ss.
XX	
OS	Synthetic.
XX	
PN	WO200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	
PI	Umek RM;
XX	
DR	WPI; 2001-159728/16.
XX	
PT	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface.
XX	
PS	Example 6; Page 128; 159pp; English.
XX	
CC	The present invention relates to a composition comprising two nucleic
CC	acids each containing an electron-transfer group (ETW) having

```
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

      Query Match          6.0%; Score 83; DB 22; Length 936;
      Best Local Similarity 1.2%; Pred. No. 7.4e-14;
      Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps
      :
Qy 482 aaactagaagtgcgagtcgaattcaattaggcgcgcaaatcacagaattcggaaatag 541
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 750 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 542 ctctcgataaccacgaggtattctgactcttaggtaaaactgacttccctgcaccttac 601
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 690 GWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 602 agactgtctctccaattctagcaacaataaacagcagctgagctctttaagaga 661
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 630 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 662 tgcagaatacccagtagtcccgaggaaacccctgcgaattgtccaatctttagtgatc 721
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 570 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 722 agacagatgctacgcgcacacagatagagaagatggaaaatcgcatgggggatcat 781
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 510 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 782 ttgcaggacacagcgtagtgagctgtagaaaaatgctaataacagtatataagca 841
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 842 acatagattcagctaagcagcgaatgcgtactgctaaagacacaaatagctgaagctcaga 901
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 390 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 902 aaaaftccccgactctcaaattctcaagaagcggaacaaatggtaatacacaggtcaga 961
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 330 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 962 aagatctttaaataatcaaacctgcagatggttctgatgttcaccaatccaggaaclacag 1021
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 1022 ttggaggctccaagacaagaagtagtatggtagtattcggttccatgctgttag 1081
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 210 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 1082 atgatgtcaaaatgagacgcgtccattttgatgtctgggttcgtcagatgatgccaca 1141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW
Qy 1142 tgttcaatacggaaaatcctgattctcaa 1170
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW WWWWWWWW

RESULT 12
AAF58255/c
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT
TT
XX
XX Deletion
DE Oligonucleotide D1876.
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
```



XX OS Synthetic.  
 XX PN WO200107665-A2.  
 XX PD 01-FEB-2001.  
 XX PF 26-JUL-2000; 2000WO-US20476.  
 XX PR 26-JUL-1999; 99US-0145695.  
 XX PR 17-MAR-2000; 2000US-0190259.  
 XX PA (CLIN-) CLINICAL MICRO SENSORS INC.  
 XX PI Umek RM;  
 XX DR WPI; 2001-159728/16.  
 XX PT Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX PS Example 6; Page 127; 159pp; English.  
 XX CC The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 6.08; Score 83; DB 22; Length 938;  
 Best Local Similarity 1.2%; Pred. No. 7.4e-14;  
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

Qy 482 aaactaagaatgcgagatcaattaaagtggcgcaaatcacagaattagcgaataatg 541  
 Db 750 www 691  
 Qy 542 ctccgataacacagcgactcttagctcttagtgtaaaactgactcttcgacctcttac 601  
 Db 690 Gwww 631  
 Qy 602 agactgctcttccaatctgacaaacataacaaagcagctgagctcttctaaagaga 661  
 Db 630 www 571  
 Qy 662 tgcaagataaccagtagtccagggaaacgcctgcaattgctcaattcttagttgac 721  
 Db 570 www 511  
 Qy 722 agacagatgctacagcgacacagatagagaaagatggaaatgcgattgggagatcatatt 781  
 Db 510 www 451  
 Qy 782 ttgcaggacagacgctagtgagctgtagaaaatgctaaaatctaaacagatataagca 841  
 Db 450 www 391  
 Qy 842 acatagattcagctaaacagcagcaatcgctactgctctgctaaacacaaatagctgaagctcaga 901  
 Db 390 www 331  
 Qy 902 aaaagttcccgactctccaattctcaagagcggaaacaaatggtataacagcgtcaga 961  
 Db 330 www 271  
 Qy 962 aagattctaaatatcaaacctcgacatggttctgatttccaaatccaggaaactacag 1021  
 Db 270 www 211

Qy 1022 ttgaggctccaagcaaggaagtagtattgtagtattgctgtttccatgctgttag 1081  
 Db 210 www 151  
 Qy 1082 atgatgctgaaatgagaccgcttcatttgatgctgggttctgcagatgattcaca 1141  
 Db 150 www 91  
 Qy 1142 tgttcaatacggaaaaatcctgattctcaa 1170  
 Db 90 ww 62

RESULT 13

AAF58252  
 ID AAF58252 standard; DNA; 936 BP.  
 XX AC AAF58252;  
 XX DT 24-APR-2001 (first entry)  
 XX DE Oligonucleotide D1835.  
 XX KW Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 XX OS Synthetic.  
 XX PN WO200107665-A2.  
 XX PD 01-FEB-2001.  
 XX PF 26-JUL-2000; 2000WO-US20476.  
 XX PR 26-JUL-1999; 99US-0145695.  
 XX PR 17-MAR-2000; 2000US-0190259.  
 XX PA (CLIN-) CLINICAL MICRO SENSORS INC.  
 XX PI Umek RM;  
 XX DR WPI; 2001-159728/16.  
 XX PT Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX PS Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.  
 Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 5.8%; Score 80.6; DB 22; Length 936;  
 Best Local Similarity 1.2%; Pred. No. 3.7e-13;  
 Matches 8; Conservative 402; Mismatches 281; Indels 0; Gaps 0;

Qy 483 aactaagaatgccgagatgcaattaaagtggcgcaaatcacagaattagcgaataatgc 542  
 Db 1 www 60  
 Qy 543 ttccgataacacagcgactcttagctcttagtgtaaaactgactcttcgacctcttaca 602  
 Db 61 www 120  
 Qy 603 gactgctcttccaatctgtagcaacaataacaaagcagctgagcttcttaagagat 662



```

Db 121 ..... 180
Qy 663 gcaagataaccacagtagtccacagggaaacacgctgcaattgctcaattttagtgatca 722
Db 181 ..... 240
Qy 723 gacagatgctacagcgacacagatagagaaagatggaaatgcgattgggagtcatttt 782
Db 241 ..... 300
Qy 783 tgcaggacagacgctagtgagctgtagaataatgctaaatctataacagatataagcaa 842
Db 301 ..... 360
Qy 843 catagattcagctaaagcagcaatgctactgctaaagacacacaaatagctgaagctcagaa 902
Db 361 ..... 420
Qy 903 aaagtcccgactctccaattcttcaagaagcggaacaaatggttaacacagctgagaa 962
Db 421 ..... 480
Qy 963 agatctataaatacaaacctgcagatggttctgatttccaaatccaggaactacagt 1022
Db 481 ..... 540
Qy 1023 tggagctccaaagcagcaagatgattggttagttgctgatttccatgctgttaga 1082
Db 541 ..... 600
Qy 1083 tgatgctgaaatgagacgcttccatttctgattgctggttctgctgagatgcacat 1142
Db 601 ..... 660
Qy 1143 gtccaatcggaaatcctgatttccagct 1173
Db 661 ..... 691

```

RESULT 14

AAF58254  
ID AAF58254 standard; DNA; 936 BP.

AC AAF58254;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1875.

KW Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

XX 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface

```

XX Example 6; Page 127; 159pp; English.
PS
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

```

```

Query Match 5.8%; Score 80.6; DB 22; Length 936;
Best Local Similarity 1.2%; Pred. No. 3.7e-13;
Matches 8; Conservative 402; Mismatches 281; Indels 0; Gaps 0;
Qy 483 aactaagaatgccgatgcaattaaagtggcgcgcaaatattacagaatagcgaatgc 542
Db 1 ..... 60
Qy 543 ttcggataaacacgagcttctgactctttaggttaaaactgacttccttcgacctttaca 602
Db 61 ..... 120
Qy 603 gactgctcttccaaatctgtagcaacaataacaaagcagctgagcttcttaagagat 662
Db 121 ..... 180
Qy 663 gcaagataaacacagtagtccacagggaaacccctgcattgctcaattttagtgatca 722
Db 181 ..... 240
Qy 723 gacagatgctacagcgacacagatagagaaagatggaaatcgattgggagtcatttt 782
Db 241 ..... 300
Qy 783 tgcaggacagacgctagtgagctgtagaataatgctaaatctataacagatataagcaa 842
Db 301 ..... 360
Qy 843 catagattcagctaaagcagcaatgctactgctaaagacacacaaatagctgaagctcagaa 902
Db 361 ..... 420
Qy 903 aaagtcccgactctccaattcttcaagaagcggaacaaatggttaacacagctgagaa 962
Db 421 ..... 480
Qy 963 agatctataaatacaaacctgcagatggttctgatttccaaatccaggaactacagt 1022
Db 481 ..... 540
Qy 1023 tggagctccaaagcagcaagatgattggttagttgctgatttccatgctgttaga 1082
Db 541 ..... 600
Qy 1083 tgatgctgaaatgagacgcttccatttctgattgctggttctgctgagatgcacat 1142
Db 601 ..... 660
Qy 1143 gtccaatcggaaatcctgatttccagct 1173
Db 661 ..... 691

```

RESULT 15

AAF58257  
ID AAF58257 standard; DNA; 936 BP.

AC AAF58257;

DT 24-APR-2001 (first entry)











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:53:51 ; Search time 243.49 Seconds  
(without alignments)  
1291.954 Million cell updates/sec

Title: US-09-391-606-4

Perfect score: 1389

Sequence: 1 atggttaactctattgttc.....aagttatgatccagact 1389

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/1na/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/1na/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146.8	10.6	150	4	US-08-659-473-8
2	54.4	3.9	7218	1	US-08-232-463-14
3	50.6	3.6	2085	1	US-08-072-070-1
4	50.6	3.6	2085	1	US-08-465-746-1
5	50.6	3.6	2085	1	US-08-214-164-1
6	50.6	3.6	2085	1	US-08-469-434-1
7	50.6	3.6	2085	1	US-08-214-222-1
8	50.6	3.6	2085	2	US-08-467-852A-1
9	50.6	3.6	2085	2	US-08-468-718-1
10	50.6	3.6	2085	2	US-08-246-636-1
11	50.6	3.6	2085	2	US-08-247-491A-1
12	50.6	3.6	2085	2	US-08-319-795-1
13	50.6	3.6	2085	2	US-08-468-985-1
14	50.6	3.6	2086	3	US-08-312-949-1
15	50.6	3.6	2086	3	US-08-446-201-2
16	43.6	3.1	6744	1	US-08-119-125A-2
17	39.6	2.9	640	3	US-08-961-083-33
18	39.4	2.8	3337	1	US-08-072-610-1
19	39.4	2.8	3337	2	US-08-719-822B-1
20	39.4	2.8	3337	4	US-09-092-458-1
21	37.4	2.7	1920	1	US-08-186-222-1
22	37.4	2.7	5361	4	US-08-973-462-2
23	37.4	2.7	5820	4	US-09-029-213B-7
24	37.4	2.7	6152	4	US-08-973-462-1
25	36.4	2.6	5738	1	US-08-409-995-3
26	36.4	2.6	5738	3	US-08-685-467-3
27	36.4	2.6	7291	4	US-08-913-942-3

28 35.6 2.6 2943 2 US-08-788-892-1 Sequence 1, Appli  
29 35 2.5 697 6 5171843-10 Patent No. 5171843  
30 35 2.5 1137 6 5171843-8 Patent No. 5171843  
31 35 2.5 1855 3 US-08-961-083-71 Sequence 71, Appli  
32 34.8 2.5 5177 6 5352450-1 Patent No. 5352450  
33 34.6 2.5 414 2 US-08-630-822A-63 Sequence 63, Appli  
34 34.6 2.5 414 2 US-09-005-069-63 Sequence 63, Appli  
35 34.6 2.5 650 4 US-08-998-416-190 Sequence 190, Appli  
36 34.6 2.5 708 4 US-09-285-576-1 Sequence 1, Appli  
37 34.6 2.5 716 4 US-08-998-416-1032 Sequence 1032, Ap  
38 34.6 2.5 835 4 US-08-998-416-533 Sequence 533, App  
39 34.6 2.5 836 4 US-08-998-416-286 Sequence 286, App  
40 34 2.4 168 1 US-08-469-802B-4 Sequence 4, Appli  
41 34 2.4 168 2 US-08-267-803B-4 Sequence 4, Appli  
42 34 2.4 171 1 US-08-469-802B-5 Sequence 5, Appli  
43 34 2.4 171 2 US-08-267-803B-5 Sequence 5, Appli  
44 34 2.4 195 1 US-08-469-802B-2 Sequence 2, Appli  
45 34 2.4 195 2 US-08-267-803B-2 Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-659-473-8  
; Sequence 8, Application US/08659473  
; Patent No. 6210876  
; GENERAL INFORMATION:  
; APPLICANT: M. B. Cerney  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.0.1  
; SOFTWARE: Microsoft Word 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/659,473  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul D. Yasger  
; REGISTRATION NUMBER: 37,477  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 708/937-2341  
; TELEFAX: 708/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 150 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA (C. pneumoniae)  
US-08-659-473-8

Query Match 10.6%; Score 146.8; DB 4; Length 150;  
Best Local Similarity 98.7%; Pred. No. 9.8e-35;  
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Oy 366 tacctcaacatcactagctgacatcacaggtcttggagcctccaggtgctgac 425  
|||||  
Db 1 TACCTCAACATCAGTACGTGACGTACAGGCTCTTTGGTGACCTCCAGGATGCTGTAC 60



Qy 426 taataaagatacagcgctactgtatgagaaacgcgaatcgtcgagtgaggaaac 485  
|||||  
Db 61 TAATATAAGGATACAGCGGTACTGTATGAGGAAACCGCAATCGTGGGTGGGAAAC 120  
|||||  
Qy 486 taagatccgctgaataaagtggcgc 515  
|||||  
Db 121 TAAGATGCCGATGCGAGTTAAGTTGGGCGC 150  
|||||

## RESULT 2

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935.313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 3.9%; Score 54.4; DB 1; Length 7218;  
Best Local Similarity 9.3%; Pred. No. 4.6e-06;  
Matches 43; Conservative 218; Mismatches 199; Indels 0; Gaps 0;

Qy 631 aatacaaacagctgagcttctaagagatgcaagataaccacagtagtccagggaa 690  
|||||  
Db 1481 AATTACTATCTATGCAAGTAGTTAAGAGATAGAGAAATTTGGTACRRRRRRRRR 1422  
|||||  
Qy 691 acgcctgaattgctcaatttttagttgatcacagatgtctacagcacacagatag 750  
|||||  
Db 1421 RRR 1362  
|||||  
Qy 751 aagagtggaatcgattggggatgcatatttgcaggacagacgctagtgagctgta 810  
|||||

Db 1361 RRR 1302  
|||||  
Qy 811 gaaaatgctaaatctaataacagataagcaacatagattcagctaaagcagcaatcgct 870  
|||||  
Db 1301 RRR 1242  
|||||  
Qy 871 acLgctaagacacaaatagctgaagctcagaaaaagtcccccagctctccaattctcaa 930  
|||||  
Db 1241 RRR 1182  
|||||  
Qy 931 gaagcggaacaaatgtaatacaggtgagaagatcttaaaaaatacaaacctgcgat 990  
|||||  
Db 1181 RRR 1122  
|||||  
Qy 991 gttctgatgttccaaatccaggaactacagtggaggctcccaagcaacaaggaagtagt 1050  
|||||  
Db 1121 RRR 1062  
|||||  
Qy 1051 attggtagtagtctgtttccatgctgttagatgtagtctg 1090  
|||||  
Db 1061 AAGCTCCCTCGACCTGCAGCAAGCTCGGAATTAATTCTG 1022  
|||||

## RESULT 3

US-08-072-070-1  
; Sequence 1, Application US/08072070  
; Patent No. 5476929  
; GENERAL INFORMATION:  
; APPLICANT: Briles, David E  
; APPLICANT: Yother, Janet L  
; APPLICANT: McDaniel, Larry S  
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL  
; TITLE OF INVENTION: PROTEIN  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: Suite 1203, 2001 Jefferson Davis Highway  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/072.070  
; FILING DATE: 19930603  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/835.698  
; FILING DATE: 12-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/656.773  
; FILING DATE: 15-FEB-1991  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 521-0378  
; TELEX: LUKPAT WASHINGTON  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2085 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; STRAIN: Rsl  
; IMMEDIATE SOURCE:



CLONE: JY2008

FEATURE:

NAME/KEY: intron

LOCATION: 1..2085

FEATURE:

NAME/KEY: CDS

LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011

LOCATION: ..2025, 2029..2031, 2035..2085)

US-08-072-070-1

Query Match

Best Local Similarity 3.6%; Score 50.6; DB 1; Length 2085;

Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

Qy 745 atagagaagaatgcgaatgcgattgggagtcataatttcgcagacagaaacgctagtga 804

Db 472 ACAGACAAGCGCAAGAGCCAGCAGATAGATGATAGTGAAGCTTAAGAACGCCAA 531

Qy 805 gctgtagaaaatgctaaatcttaataacagtagtaagcaacatagattcagctaaagcagca 864

Db 532 GAAGAGGCAAAACCTAAATTTAACTGTCGAGCAATGGTAGTCTCTGAGCCAGCAG 591

Qy 865 atcgctactgtaagacaacaaatagctgaagctcagaaagagttcccca ---ctctcca 921

Db 592 TTGGCTGAGACTAAGAAAAATCAAGAAAGCTTAACCAAAAGCACCAGAACTTACTAAA 651

Qy 922 attctcagaagcggaacaaatgtaatacagcgctgagaaagatcttaaaaaatatcaaa 981

Db 652 AAAGTAGAAGAGCTTAAGCAAAATTAAGAGAGGCTGAGAAAAAGCTTACTGAAGCCAAA 711

Qy 982 cctgcagatggtctgtatgttccaaatccaggaactacagttggaggctcccaagcaaaa 1041

Db 712 CAAAAAGTGGATGCTGAAGAGTCTCTCTCAAGCTAAATCGCTGAATTTGAAAAATCAA 771

Qy 1042 ggaagtagtattg 1054

Db 772 GTTCATAGACTAG 784

RESULT 4

US-08-465-746-1

Sequence 1, Application US/08465746

Patent No. 5679768

GENERAL INFORMATION:

APPLICANT: Briles, David E

APPLICANT: Yother, Janet L

APPLICANT: McDaniel, Larry S

TITLE OF INVENTION: EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE

TITLE OF INVENTION: PROTEIN A

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: Suite 1203, 2001 Jefferson Davis Highway

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/465.746

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/048.896

FILING DATE:

APPLICATION NUMBER: US 07/656.773

FILING DATE: 15-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/835.698

FILING DATE: 12-FEB-1992

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 521-0378

TELEX: LUKPAT WASHINGTON

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2085 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

STRAIN: RX1

IMMEDIATE SOURCE:

CLONE: JY4313

FEATURE:

NAME/KEY: intron

LOCATION: 1..2085

FEATURE:

NAME/KEY: CDS

LOCATION: join(127..1984)

US-08-465-746-1

Query Match

Best Local Similarity 3.6%; Score 50.6; DB 1; Length 2085;

Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

Qy 745 atagagaagaatgcgaatgcgattgggagtcataatttcgcagacagaaacgctagtga 804

Db 472 ACAGACAAGCGCAAGAGCCAGCAGATAGATGATAGTGAAGCTTAAGAACGCCAA 531

Qy 805 gctgtagaaaatgctaaatcttaataacagtagtaagcaacatagattcagctaaagcagca 864

Db 532 GAAGAGGCAAAACCTAAATTTAACTGTCGAGCAATGGTAGTCTCTGAGCCAGCAG 591

Qy 865 atcgctactgtaagacaacaaatagctgaagctcagaaagagttcccca ---ctctcca 921

Db 592 TTGGCTGAGACTAAGAAAAATCAAGAAAGCTTAACCAAAAGCACCAGAACTTACTAAA 651

Qy 922 attctcagaagcggaacaaatgtaatacagcgctgagaaagatcttaaaaaatatcaaa 981

Db 652 AAAGTAGAAGAGCTTAAGCAAAATTAAGAGAGGCTGAGAAAAAGCTTACTGAAGCCAAA 711

Qy 982 cctgcagatggtctgtatgttccaaatccaggaactacagttggaggctcccaagcaaaa 1041

Db 712 CAAAAAGTGGATGCTGAAGAGTCTCTCTCAAGCTAAATCGCTGAATTTGAAAAATCAA 771

Qy 1042 ggaagtagtattg 1054

Db 772 GTTCATAGACTAG 784

RESULT 5

US-08-214-164-1

Sequence 1, Application US/08214164

Patent No. 5728387

GENERAL INFORMATION:

APPLICANT: Briles, David E.

APPLICANT: YOTHER, JANET L.

TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: Suite 1203, 2001 Jefferson Davis Highway

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.



ZIP: 22202-0286  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/214,164  
 FILING DATE: 17-MAR-1994  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/656,773  
 FILING DATE: 15-FEB-1991  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Berkstresser, Jerry W.  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: 6102-137  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 521-0813  
 TELEX: LUKPAT WASHINGTON  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2085 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 1..1983  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 127..1983  
 US-08-214-164-1

Query Match 3.6%; Score 50.6; DB 1; Length 2085;  
 Best Local Similarity 49.8%; Pred. No. 3.1e-05;  
 Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;  
 QY 745 atagagaaagatggaatcgattggggtgcatattttgcaggacacagaacgctagtggga 804  
 Db 472 ACAGACAAAGCGCGAAAGACGACGACGATAGATGATAGATGAAGCTAAGAAAGCGCAA 531  
 QY 805 gctgtagaaatgctaaatcttaataacaggtataagcaacatagattcagctaaagcagca 864  
 Db 532 GAAGAGGCAAAACTAAATTTAATACTGTCGAGCAATGGTAGTTCCTGAGCCAGAGCAG 591  
 QY 865 atcgctactgtaagacacaaatagctgaagctcagaaaagttcccca---ctctcca 921  
 Db 592 TTGGCTGAGACTAGAAAATAATCAGAAGAGCTAACAATAAGACACAGAACTTACTTAA 651  
 QY 922 attctcagaagcgggaacaaatgtaatacagggctgagaagatctttaaataatcaaa 981  
 Db 652 AAACATAGAAGAGCTAAGCAAAATTAGAAGAGGCTGAGAAAAGCTACTGAAGCCAAA 711  
 QY 982 cctgcagatggtctgattgttcccaatccaggaaactacagttggagggtccaaagcaaaa 1041  
 Db 712 CAAAAGTGGATGCTGAAGAGAGTCTCTCTCAAGCTAAATCGCTGAATGGAAAATCAA 771  
 QY 1042 ggaagtagattg 1054  
 Db 772 GTTCATAGACTAG 784  
 RESULT 6  
 US-08-469-434-1  
 ; Sequence 1, Application US/08469434  
 ; Patent No. 5753463  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Briles, David E

APPLICANT: Yother, Janet L  
 APPLICANT: McDaniel, Larry S  
 TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL  
 TITLE OF INVENTION: PROTEIN  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Mattare, Ltd  
 STREET: Suite 1203, 2001 Jefferson Davis Highway  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/469,434  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/072,065  
 FILING DATE: 03 JUNE 1993  
 APPLICATION NUMBER: US/07/835,698  
 FILING DATE: 12-FEB-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/656,773  
 FILING DATE: 15-FEB-1991  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 521-0378  
 TELEX: LUKPAT WASHINGTON  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2085 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Streptococcus pneumoniae  
 STRAIN: Rx1  
 IMMEDIATE SOURCE:  
 CLONE: JY2008  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 1..2085  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011  
 LOCATION: ..2025, 2029..2031, 2035..2085)  
 US-08-469-434-1

Query Match 3.6%; Score 50.6; DB 1; Length 2085;  
 Best Local Similarity 49.8%; Pred. No. 3.1e-05;  
 Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;  
 QY 745 atagagaaagatggaatcgattggggtgcatattttgcaggacacagaacgctagtggga 804  
 Db 472 ACAGACAAAGCGCGAAAGACGACGACGATAGATGATAGATGAAGCTAAGAAAGCGCAA 531  
 QY 805 gctgtagaaatgctaaatcttaataacaggtataagcaacatagattcagctaaagcagca 864  
 Db 532 GAAGAGGCAAAACTAAATTTAATACTGTCGAGCAATGGTAGTTCCTGAGCCAGAGCAG 591  
 QY 865 atcgctactgtaagacacaaatagctgaagctcagaaaagttcccca---ctctcca 921  
 Db 592 TTGGCTGAGACTAGAAAATAATCAGAAGAGCTAACAATAAGACACAGAACTTACTTAA 651











CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/246.636  
FILING DATE: 20-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/656,773  
FILING DATE: 15-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/835,698  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,896  
FILING DATE: 20-APR-1993  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEX: LURPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2085 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
STRAIN: Rxl  
IMMEDIATE SOURCE:  
CLONE: JY4313  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1..2085  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(127..1984)  
US-08-246-636-1

Query Match 3.6%; Score 50.6; DB 2; Length 2085;  
Best Local Similarity 49.8%; Pred. NO. 3.1e-05;  
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;  
QY 745 atagagaagaatggaatgcgattgggagtcataattttgcaggacagaaacgctagtga 804  
DB 472 ACAGACAAGCCGCAAAAGACCCAGCAGATAAGATGATAGATGAAGCTAAGAAACCGCAA 531  
QY 805 gctgtagaaaatgctaaatcttaataacagataaagcaacatagattcagctaaagcagca 864  
DB 532 GAAGAGGCAAAACCTAAATTTAATCTGTTCCAGCAATGGTAGTCTTCCAGCCAGCAG 591  
QY 865 atcgctactgctaaagacacaaatagctgaagctcagaaagttccccga---ctctcca 921  
DB 592 TTGGCTGAGACTTAAGAAAAATCAGAAGAGCTTAACAAAAAGCACCAGAACTTACTAAA 651  
QY 922 attcttcaagagcggaacaaatgtaatacaggtcgagaaagcttcaaaatcaaa 981  
DB 652 AAACATAGAAAGCTTAAGCAAAATTAAGAGAGCTGAGAAAAAGCTACTGAAGCCAAA 711  
QY 982 cctgcagatgggtctgattgtcccaatccaggaaactacagttggaggctcccaagcaaaa 1041  
DB 712 CAANAAGTGGATGCTGAAGAAGTCCTCTCAAGCTAAATCGTGAATGGAAAAATCAA 771

QY 1042 ggaagtagtattg 1054  
DB 772 GTTCATAGACTAG 784  
RESULT 11  
US-08-247-491A-1  
Sequence 1, Application US/08247491A  
Patent No. 5965400  
GENERAL INFORMATION:  
APPLICANT: BRILES, David E.  
APPLICANT: YOTHER, Janet L.  
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
STREET: 745 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10151  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/247,491A  
FILING DATE: 23-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOWALSKI, Thomas J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454312-2041  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2085 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
STRAIN: Streptococcus pneumoniae Rxl  
IMMEDIATE SOURCE:  
CLONE: JY2008  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1..2085  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011  
LOCATION: ..2025, 2029..2031, 2035..2085)  
US-08-247-491A-1  
Query Match 3.6%; Score 50.6; DB 2; Length 2085;  
Best Local Similarity 49.8%; Pred. NO. 3.1e-05;  
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;  
QY 745 atagagaagaatggaatgcgattgggagtcataattttgcaggacagaaacgctagtga 804  
DB 472 ACAGACAAGCCGCAAAAGACCCAGCAGATAAGATGATAGATGAAGCTAAGAAACCGCAA 531  
QY 805 gctgtagaaaatgctaaatcttaataacagataaagcaacatagattcagctaaagcagca 864  
DB 532 GAAGAGGCAAAACCTAAATTTAATCTGTTCCAGCAATGGTAGTCTTCCAGCCAGCAG 591  
QY 865 atcgctactgctaaagacacaaatagctgaagctcagaaagttccccga---ctctcca 921



Db 592 TTGGCTGAGACTAAGAAAATCAGAGAGCTAAACAAAAGCACCAGAACTTACTTAAA 651  
Qy 922 attcttcaagcggcaaatggttaatacagcgtgagaaagatcttataaaatatcaaa 981  
Db 652 AAATAGAGAGAGCTAAGCAAAATTTAGAGAGGCTGAGAAAAGGCTACTGAAAGCCAAA 711  
Qy 982 cctgcagatggttctgattccaaatccaggaaactacagttgagggtccaaagcaaaa 1041  
Db 712 CAAAAGTGGATGCTGAAGAAGTCCTCTCAAGCTAAATTCGTGAATTCGAAAATCAA 771  
Qy 1042 ggaagtagtattg 1054  
Db 772 GTTCATAGACTAG 784

## RESULT 12

US-08-319-795-1  
; Sequence 1, Application US/08319795  
; Patent No. 5980909

## GENERAL INFORMATION:

APPLICANT: Briles, David E.  
APPLICANT: Yother, Janet L.  
APPLICANT: McDaniel, Larry S.  
TITLE OF INVENTION: Epitopic Regions of Pneumococcal Surface  
TITLE OF INVENTION: Protein A  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheomaker and Mattare, Ltd.  
STREET: 1203 Crystal Plaza Bldg. 1, 2001 Jefferson  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/319,795  
FILING DATE:

## CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/246,636  
FILING DATE: 20-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,896  
FILING DATE: 20-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/835,698  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/656,773  
FILING DATE: 15-FEB-1991  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2085 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
STRAIN: Rx1  
IMMEDIATE SOURCE:  
CLONE: JY4313

## FEATURE:

NAME/KEY: Intron  
LOCATION: 1..2085  
FEATURE:  
NAME/KEY: CDS  
LOCATION: Join(127..1984)  
US-08-319-795-1

## Query Match 3.6%; Score 50.6; DB 2; Length 2085;

Best Local Similarity 49.8%; Pred. No. 3.1e-05;  
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

Qy 745 atagagaagaatggaatgcatgattggggatgcatttttcagagacagacgcctagtggga 804  
Db 472 ACAGACAAGCCGCAAAAGACGACAGATAAGATGATAGATGAAGCTAAGAAACGCGAA 531  
Qy 805 gctgtagaaaatgctaaatcttaataacagtcataataagcaacatagattcagctaaagcagca 864  
Db 592 TTGGCTGAGACTAAGAAAATCAGAGAGCTAAACAAAAGCACCAGAACTTACTTAAA 651  
Qy 922 attcttcaagcggcaaatggttaatacagcgtgagaaagatcttataaaatatcaaa 981  
Db 652 AAATAGAGAGAGCTAAGCAAAATTTAGAGAGGCTGAGAAAAGGCTACTGAAAGCCAAA 711  
Qy 982 cctgcagatggttctgattccaaatccaggaaactacagttgagggtccaaagcaaaa 1041  
Db 712 CAAAAGTGGATGCTGAAGAAGTCCTCTCAAGCTAAATTCGTGAATTCGAAAATCAA 771  
Qy 1042 ggaagtagtattg 1054  
Db 772 GTTCATAGACTAG 784

## RESULT 13

US-08-468-985-1  
; Sequence 1, Application US/08468985  
; Patent No. 5997882

## GENERAL INFORMATION:

APPLICANT: Briles, David E.  
APPLICANT: Yother, Janet L.  
APPLICANT: McDaniel, Larry S.  
TITLE OF INVENTION: Epitopic Regions of Pneumococcal Surface  
TITLE OF INVENTION: Protein A  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheomaker and Mattare, Ltd.  
STREET: 1203 Crystal Plaza Bldg. 1, 2001 Jefferson  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,985  
FILING DATE:

## CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/319,795  
FILING DATE:  
APPLICATION NUMBER: US 08/246,636  
FILING DATE: 20-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,896



FILING DATE: 20-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/835,698  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/656,773  
FILING DATE: 15-FEB-1991  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2085 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
STRAIN: Rx1  
IMMEDIATE SOURCE:  
CLONE: JY4313  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1..2085  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(127..1984)  
US-08-468-985-1

Query Match 3.6%; Score 50.6; DB 2; Length 2085;  
Best Local Similarity 49.8%; Pred. No. 3.le-05;  
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;  
QY 745 atagagaagaatggaatgcgattgggagtcataatttgcaggacagaacgctagtga 804  
DB 472 ACAGACAAAGCGCGCAAAAGACGACGACGATAGATGATAGATGAAGCTAAGAAACGCGAA 531  
QY 805 gctgtagaaaatgctaaatcctaataacagataaagcaacatagattcagctaaagcagca 864  
DB 532 GAAGAGGCAAAACCTAAATTTAATACTGTCGAGCAATGGTATCTCTCCTGAGCAGCAG 591  
QY 865 atcgtactgctaaagacacaaatagctgaagctcagaaaaagttcccca---ctctcca 921  
DB 592 TTGGCTGAGACTAAGAAAAATCAGAAAGCTTAACAAAAAGCAGCACTTACTTAA 651  
QY 922 attctcagaagcggaacaaatgtaatacaggtcgagaagatcttataaataatcaaa 981  
DB 652 AAACCTAGAAGAGCTTAAAGCAAAATTAGAAGAGGCTGAGAAAAAGCTACTGAAGCCAAA 711  
QY 982 cctgcagatggtctgattgttcccaatccaggaaactacagttggagggtcccaagcaaaa 1041  
DB 712 CAAAAGTGGATGCTGAAGAAAGTCGCTCTCAAGCTAAATCGCTGAATTTGAAAAATCAA 771  
QY 1042 ggaagtagtattg 1054  
DB 772 GTTCATAGACTAG 784

RESULT 14  
US-08-312-949-1  
Sequence 1, Application US/08312949  
Patent No. 6027734  
GENERAL INFORMATION:  
APPLICANT: Briles, David E.  
APPLICANT: Wu, Hong-Yin  
TITLE OF INVENTION: MUCOSAL ADMINISTRATION OF  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/312.949  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454312-2049  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2086 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-312-949-1

Query Match 3.6%; Score 50.6; DB 3; Length 2086;  
Best Local Similarity 49.8%; Pred. No. 3.le-05;  
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;  
QY 745 atagagaagaatggaatgcgattgggagtcataatttgcaggacagaacgctagtga 804  
DB 472 ACAGACAAAGCGCGCAAAAGACGACGACGATAGATGATAGATGAAGCTAAGAAACGCGAA 531  
QY 805 gctgtagaaaatgctaaatcctaataacagataaagcaacatagattcagctaaagcagca 864  
DB 532 GAAGAGGCAAAACCTAAATTTAATACTGTCGAGCAATGGTATCTCTCCTGAGCAGCAG 591  
QY 865 atcgtactgctaaagacacaaatagctgaagctcagaaaaagttcccca---ctctcca 921  
DB 592 TTGGCTGAGACTAAGAAAAATCAGAAAGCTTAACAAAAAGCAGCACTTACTTAA 651  
QY 922 attctcagaagcggaacaaatgtaatacaggtcgagaagatcttataaataatcaaa 981  
DB 652 AAACCTAGAAGAGCTTAAAGCAAAATTAGAAGAGGCTGAGAAAAAGCTACTGAAGCCAAA 711  
QY 982 cctgcagatggtctgattgttcccaatccaggaaactacagttggagggtcccaagcaaaa 1041  
DB 712 CAAAAGTGGATGCTGAAGAAAGTCGCTCTCAAGCTAAATCGCTGAATTTGAAAAATCAA 771  
QY 1042 ggaagtagtattg 1054  
DB 772 GTTCATAGACTAG 784

RESULT 15  
US-08-446-201-2  
Sequence 2, Application US/08446201B  
Patent No. 6042838  
GENERAL INFORMATION:  
APPLICANT: Briles, David E.  
APPLICANT: Wu, Hong-Yin  
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR MUCOSAL ADMINISTRATION OF  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
CURRENT FILING DATE: 1995-05-19



```

; EARLIER APPLICATION NUMBER: 08/312,949
; EARLIER FILING DATE: 1994-09-30
; EARLIER APPLICATION NUMBER: 08/246,636
; EARLIER FILING DATE: 1994-05-20
; EARLIER APPLICATION NUMBER: 08/048,896
; EARLIER FILING DATE: 1993-04-20
; EARLIER APPLICATION NUMBER: 07/835,698
; EARLIER FILING DATE: 1992-02-12
; EARLIER APPLICATION NUMBER: 07/656,773
; EARLIER FILING DATE: 1991-02-15
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2086
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-08-446-201-2

```

```

Query Match      3.6%; Score 50.6; DB 3; Length 2086;
Best Local Similarity 49.8%; Pred. No. 3.1e-05;
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

QY 745 atagagaagatggaatgcgattggggtgcatattttgcagacagacgctagtggg 804
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 acagacaaagccgcaaaagcagcagcagcagcagcagcagcagcagcagcagc 531
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 805 gctgtagaaaatgctaaatctaataaacagtagtaagcaacatagattcagctaaagcagca 864
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 gaagaggcaaaaactaaatttaactgttcgagcaatggtagttcctgagccagagcag 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 865 atcgctactgctaaagacaaatagctgaagctcagaaaaagttcccca---ctctcca 921
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 ttggctgagactaaagaaaaatcagaagaagcctaataacaaaagcaccagaaacttaataa 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 922 attctcaagaagcgggaacaaatggttaatacaggtgctgagaagatcttataaaatatcaaa 981
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 aaactagaagaagcctaaagcaaaattagaaggggctgagaaaaagctactgaagcccaaa 711
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 982 cctgcagatggttcgtgattcccaaatccaggaactacagttggaggctcccaagcaacaa 1041
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 caaaaagtggatgctgagaagtcgctcctcctaagctaaatcgctgaattggaaaatcaa 771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1042 ggaagtagtatgtg 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 gtccatagactag 784
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: February 7, 2002, 15:54:07  
Job time: 361 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 21:32:23 ; Search time 7419.31 seconds  
(without alignments)  
2011.762 Million cell updates/sec

Title: US-09-391-606-4  
Perfect score: 1389  
Sequence: 1 atgtgtaacccattgtgtcc.....aagttctatggatcgagct 1389

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST: \*
- 1: em\_estfun: \*
  - 2: em\_esthum: \*
  - 3: em\_estin: \*
  - 4: em\_estom: \*
  - 5: em\_estpl: \*
  - 6: em\_estba: \*
  - 7: em\_estro: \*
  - 8: em\_estov: \*
  - 9: em\_htc: \*
  - 10: gb\_estl: \*
  - 11: gb\_est2: \*
  - 12: gb\_htc: \*
  - 13: gb\_gss: \*
  - 14: em\_gss\_fun: \*
  - 15: em\_gss\_hum: \*
  - 16: em\_gss\_inv: \*
  - 17: em\_gss\_pln: \*
  - 18: em\_gss\_pro: \*
  - 19: em\_gss\_rnd: \*
  - 20: em\_gss\_vrt: \*
  - 21: em\_gss\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	45	3.2	861	13	AZ682538 ENT1ALT
C 2	45	3.2	906	13	AZ550193 ENT254TR
C 3	44.4	3.2	484	10	AA314486 EST186534
C 4	42.8	3.1	856	13	BH132709 ENTNG59TF
C 5	42.6	3.1	460	10	AA672611 VC59B01.I
C 6	42	3.0	624	11	BF865301
C 7	41.8	3.0	586	13	AZ632588
C 8	41.8	3.0	850	10	AV406016
C 9	41.6	3.0	762	10	BE661568
C 10	41.4	3.0	842	13	AZ536150
C 11	41.4	3.0	845	13	AZ685553
C 12	41.4	3.0	876	13	AZ691916

13	41.2	3.0	616	13	AZ525704
C 14	41.2	3.0	737	10	AU075947
C 15	41.2	3.0	1000	11	BG291509
C 16	40.8	2.9	841	13	AZ676761
C 17	40.6	2.9	450	11	BG893782
C 18	40.6	2.9	527	13	AQ227560 HS_2019_B
C 19	40.6	2.9	1042	13	CNS0148K
C 20	40.4	2.9	1101	13	CNS00L4E
C 21	39.8	2.9	915	13	CNS07CXR
C 22	39.8	2.9	938	13	CNS006FJ
C 23	39.6	2.9	422	11	R15357
C 24	39.6	2.9	502	13	B60448
C 25	39.6	2.9	720	10	BE783140
C 26	39.6	2.9	946	11	BG339626
C 27	39.4	2.8	262	10	BB454826
C 28	39.4	2.8	618	11	BI394077
C 29	39.4	2.8	856	13	AZ667395
C 30	39.2	2.8	298	10	A1218089
C 31	39.2	2.8	384	11	R03975
C 32	39.2	2.8	695	10	BE217675
C 33	39.2	2.8	855	13	AZ688493
C 34	39.2	2.8	863	13	AZ548883
C 35	39.2	2.8	885	13	AZ693152
C 36	39.2	2.8	938	13	AZ549589
C 37	39	2.8	468	13	AZ522521
C 38	39	2.8	576	11	BG602389
C 39	39	2.8	748	11	BG124392
C 40	39	2.8	884	13	AZ136778
C 41	39	2.8	904	13	AZ692954
C 42	39	2.8	935	13	CNS033B4
C 43	39	2.8	1013	10	AL564621
C 44	38.8	2.8	450	13	FR0025683
C 45	38.8	2.8	738	10	AV405991

ALIGNMENTS

RESULT 1  
AZ682538/c  
LOCUS ENT1ALT  
DEFINITION ENT1ALT Entamoeba histolytica Sheared DNA  
ACCESSION AZ682538  
VERSION AZ682538.1 GI:11819684  
KEYWORDS GSS:  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica  
REFERENCE 1 (bases 1 to 861)  
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
JOURNAL HMI:IMSS sheared DNA library  
COMMENT Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: entaetigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 36  
High quality sequence stop: 813.  
Location/Qualifiers  
1. 861  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"



/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica sheared DNA"  
/note="Vector: pHOSt; Site.1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:77-7450). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Brell. Oxford University Press, 1999)."

BASE COUNT	453 a	79 c	172 g	202 t
ORIGIN				

Query Match 3.2%; Score 45; DB 13; Length 906;  
Best Local Similarity 48.3%; Pred. No. 0.36;

Qy 678 agtccacagggaaaaagccctgcgaattgctcaatcttttagttgatcagacagagatgctacagc 737

Qy 738 gacacagatagagaaaagatggaaatcgcaattgggatgcataatttgcagagacagaaacgc 797

QY 798 tagtggagctgtagaaaaatgctaaatcttaataacagtaaacacatagattcagctaa 857

Qy 858 agcagcaatcgctactgctaagacacaaatagctgaagctcagaaaagtccccgactc 917

Qy 918 tccaattcttcaagaagcgga 938

AA314486/c	LOCUS	AA314486	484 bp	mRNA	EST	19-APR-1997
DEFINITION	EST186534	Colon carcinoma (HCC) cell line II	Homo sapiens	CDNA 5'		

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A., C.J., Lee, N.H., Kirkness, E.F., Weinszock, K.G., Gocayne, J.D., White, T.R., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, I.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagan, N.S., Glodex, A.,

Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

Dimke, D., Feng, D.-F., Ferlie, A., Fischer, C., Hastings, G.A., Hewitt,  
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,  
Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,



Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse

Location/Qualifiers

#### FEATURES

source

1. .484

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):111272"

/db\_xref="taxon:9606"

/clone\_lib="Colon carcinoma (HCC) cell line II"

/tissue\_type="colon"

/cell\_type="KM12c"

/cell\_line="KM12C(HCC)-parental human colon carcinoma

;Dukes B2"

/note="Organ: colon; Vector: pBluescript SK-; Site\_1:

ECORI; Site\_2: XhoI"

BASE COUNT 133 a 182 c 107 g 61 t 1 others

#### ORIGIN

Query Match 3.2%; Score 44.4; DB 10; Length 484;

Best Local Similarity 54.2%; Pred. No. 0.43;

Matches 90; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 1171 gctgcccaacaggagctgcgcagcagcaagctagagcagcgaagcgcgtggagatgacagt 1230

Db 379 GCTGCTGGAGCTGGCGAGCTGCTGCTGGAGCTGCTGCTGGAGCTGCTGCTGGAGCT 320

Qy 1231 gctgctcagcgtgagcagctcagcagcagcagcagcagcagcagcagcagcagcagc 1290

Db 319 GCTGCTCAGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 260

Qy 1291 caacaacaggcactactcaatgcttgggacagatcgcttctgctg 1336

Db 259 GCTGTGGAGGATGCAACTGTTGCTGTGGAACCTGTAGCTGTGCTG 214

RESULT 4

BH132709/c

LOCUS ENTNG59TF Entamoeba histolytica Sheared DNA 856 bp GSS 07-AUG-2001

DEFINITION genomic, DNA sequence.

ACCESSION BH132709

VERSION BH132709.1 GI:15091760

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 856)

Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.

Determination of clone end sequences from Entamoeba histolytica

HMI:IMSS sheared DNA library (2001)

Unpublished (2001)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: enta@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: Shotgun

High quality sequence start: 3

High quality sequence stop: 856.

#### FEATURES

source

1. .856

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db\_xref="taxon:5759"

/clone\_lib="Entamoeba histolytica Sheared DNA"

/note="vector: pHOsi; Site\_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark, C.G.,

and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999).

BASE COUNT 239 a 163 c 99 g 355 t

#### ORIGIN

Query Match 3.1%; Score 42.8; DB 13; Length 856;

Best Local Similarity 46.1%; Pred. No. 1.4;

Matches 143; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 706 caactcttagttgacacagatgctacagcagcagcagcagcagcagcagcagcagcagcagc 765

Db 468 CAACTGTTACTAGTAAATAGATCTTATTTAAAGATATGCAAAAAGATCCCAAGT 409

Qy 766 attgggggagtcataatttgcaggacagcagcagcagcagcagcagcagcagcagcagc 825

Db 408 AAAAAGAAAGCTGCTGAAGCAAGTCAAGATCTTAAACAAATTTGTAAGCTACTTTACAA 349

Qy 826 aatacagcgtatacgaacacagcagcagcagcagcagcagcagcagcagcagcagcagc 885

Db 348 GAACGAGTCTTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 289

Qy 886 atagtgaagctcagaaaaagttcccgagcagcagcagcagcagcagcagcagcagcagcagc 945

Db 288 CAAGCAGAAATTAATTCACAACTCCAAACTCTGAAACTGCGAGTGGTGTTCATAGAGAT 229

Qy 946 gtatacaggctgagaaagatcttataaaatatacaaacctgcagatgggtctgatgttcca 1005

Db 228 AAAGAAATATGAGAGGAAATTAATTAAGCAGTACAGAGATGTTAAAGTACTGCAATTCAA 169

Qy 1006 aatccaggaa 1015

Db 168 TTACAACAAA 159

RESULT 5

AA672611/c

LOCUS AA672611 460 bp mRNA EST 25-NOV-1997

DEFINITION vo59h01.r1 Soares\_mammary\_gland\_NbMMG Mus musculus cDNA clone

IMAGE:1054225 5', mRNA sequence.

ACCESSION AA672611

VERSION AA672611.1 GI:2644828

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 460)

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,







Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 586)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: dunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0487 row: L column: 17  
Seq primer: CGTTGTAAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 586.

## FEATURES

source

1. 586  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0487L17"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid pL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

ORIGIN

152 a 127 c 125 g 182 t

## Query Match

Best Local Similarity

Matches

97; Conservative

0; Mismatches

92; Indels

0; Gaps

0;

3.08; Score 41.8; DB 13; Length 586;

51.3%; Pred. No. 2.2;

0; Mismatches

92; Indels

0; Gaps

0;

Qy 1011 aggaactacagttgagagctcaagcaaggaagtagtattgtagtattcgtttc 1070

Db 209 AGAGGAAGACGAGGAAGCAACAGCTAGGGCAGAGAAGACTGGTGAACCTGAAGCCA 150

Qy 1071 catgtgttagatgctgaaatgagaccgctccattttagtctggtttcgtca 1130

Db 149 AAACCCACTAGTCTCGAGAGGCCCAAAATTCGATTGCATTTTGGGCTCTGTG 90

Qy 1131 gatattcaatgttcaatcaggaataatcctgattctcaagctcccaacagggagctgc 1190

Db 89 GGTACTGGGCATGAGTAGGAGCCTAGAAAATTCGTCTCAAGACACCCCAAGACCATCTC 30

Qy 1191 agcaagc 1199

| | | | |

Db 29 AACACAAGC 21

## RESULT 8

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AV406016 850 bp mRNA EST

AV406016 Bombyx mori wing disk C108 5th-instar day-3 larva Bombyx

mori cDNA clone wdv30506 T3, mRNA sequence.

AV406016

AV406016.1 GI:6910104

EST.

Bombyx mori

domestic silkworm.

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia

; Bombycoidea; Bombycidae; Bombyx.

Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.

Bombyx mori CDNA

Unpublished (2000)

Contact: Mita, K

Genome Research Group

National Institute of Radiological Sciences

Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

Email: knita@nirs.go.jp

method: uni-directional, sequence direction: sequenced from T3 primer

(5' -&gt; 3')

Project: 'Silkworm Genome Program in MAFF, and Research for the

Future Program in JSPS'. see 'SilkBase',

&lt;http://www.ab.a.u-tokyo.ac.jp/silkbase/&gt;, for whole ESTdb.

Location/Qualifiers

1. 850

/organism="Bombyx mori"

/strain="C108"

/db\_xref="taxon:7091"

/clone="wdv30506"

/clone\_lib="Bombyx mori wing disk C108 5th-instar day-3

larva"

/sex="female/male mixed"

/tissue\_type="wing disk"

/dev\_stage="5th-instar day-3 larva"

BASE COUNT

322 a 149 c 189 g 190 t

ORIGIN

Query Match

Best Local Similarity

Matches

127; Conservative

0; Mismatches

142; Indels

0; Gaps

0;

3.08; Score 41.8; DB 10; Length 850;

47.28; Pred. No. 2.5;

0; Mismatches

142; Indels

0; Gaps

0;

Qy 697 gcaattgctcaattttagttagtcacagatgctacagcagacagatagagaaat 756

Db 459 GAATAGCTGAAGAAACCTCCCAAGAAACAGAACTACACGGGAAAGCAGAAATGAT 518

Qy 757 ggaatgcgattg99gattgcatattttgcagacagacgcgtagtg99gctgtagaaat 816

Db 519 GAACGACACACCATTTGAAACACTCTGAAAGAAAGAGAGACTAGTCTCTGTTAAAGAAAC 578

Qy 817 gctaaatctaataacagtcataagcaacatagatttcagcctaaagcagcaatcgctactgt 876

Db 579 TCACCAATTAAGAAATTTGAAGAACTACAAAAGAAAGCTACACCTGTTCTCTGTAGATAAT 638

Qy 877 aagcacacaaatagctgaagctcagaaaaagttccccgactctccaattctcaagaagcg 936

Db 639 GAAGTGCAGGAAGATGTTGCTTAACAAAGTTTCACCCGAGAGAGGAGCTGACCATCTCA 698

Qy 937 gaacacaaatggttaatacaggtcagaaaaa 965

Db 699 CAAATTTGGTTGGAGATGAAGCAAAAGA 727

RESULT 9

LOCUS

RE661568 762 bp mRNA EST

06-SEP-2000



FEATURES	Location/Qualifiers
source	1. .842

1. .842  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica sheared DNA"  
/note="vector: pHSI; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999), "

H.O. and Venter, J.C.: (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

	Query Match	3.0%	Score 41.4	DB 13	Length 842
	Best Local Similarity	48.1%	Prod. No. 3.2		
	Matches 117	Conservative	0	Mismatches 126	Indels 0
	Gaps				
Qy	820	aaatctaataacagfataagcaacatagattcagctaaagcagcaatcgctactgctaag	879		
Db	478	AAAAAATATAAANNTTCTAAAAAAGAAAGAAACAAAGAGATGGTACGACTATG	419		
Qy	880	acacaaaatagctgaagctcagaaaaagttccccgactctccaatctttccaagaagcgaa	939		
Db	418	GCACATTTAAATGAATGTATTAAATAAGTCAGTGAATTAGAAGTTATTTCATGAAGAAGTT	359		
Qy	940	caaatggtaatacaagctgagaaaatcttaaaataatacaaacctgcagatggtctgat	999		
Db	358	ATAAAGAAATTTGTACAAATTTAAAGTTATTGTTTAAAGACACAAANCTGTTTCCA	299		
Qy	1000	gttccaaatccaggaaactacagtgtggaggtctccaagcaacaaggagtagtatgtgtagt	1059		
Db	298	ATTAATAGTCACAGTAACAAATCTTCGATCTTTTAAATGGACAATTCATCATGATTTAAAT	239		

238	ATT	236
238	ATT	236







db 560 ATT 562

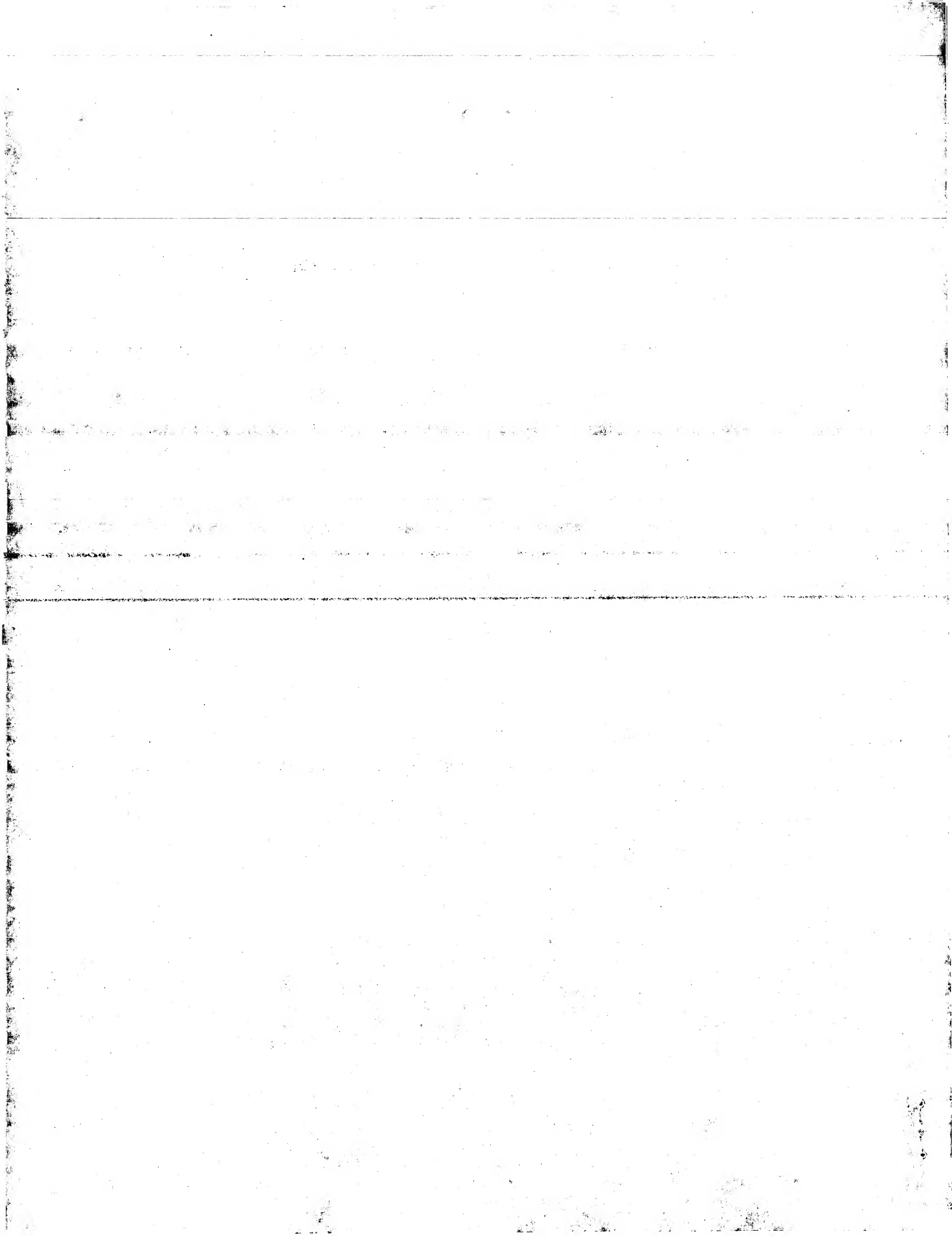
Location/Qualifiers  
1. 616

QY 774 · tgcataatttgcagacacagaacgctagtggactgtatgaaaaatgctaataaacag 833  
||| ||| | | | |||| | | |||| | |||  
pb 105. ATCATATTATGACATTAATAAAGAGTGTATTGGAGCTATCGACTTCGATTTCAGATAAAANTCATATACC 164



Search completed: February 7, 2002, 21:32:29  
Job time: 20663 sec











Db 443 THNPASEQTLQIAEALGL-----LVQEAFTDWTYGG-KKP-YDYGRFFKXATHPE 491  
 Qy 88 AKA-----TKTTL-NGKENLA---WFICGTLGLKAGDWSATVRYEVEALSVEI 135  
 Db 492 ARKGEKWSDFDLRTWVERGKNNPAIFMWSIGNEIGANGDAHSLATVK-RLVKV--IKDV 548  
 Qy 136 DVSGIGRGNLLKFWFA-----QAIAANYDPKEANSFTNYKGFSAFY---MY 178  
 Db 549 DKTRVTMGADKFRGNGSGGHEKIADELDAVGNV-----SEDNKALAKHPKWLII 602  
 Qy 179 GITSLSFRAYGAYSKP-----ANDKLG-----SDFTFRKFDLGI 213  
 Db 603 GSETSATRTGSGYRPERELKHSNGPBNVEQSDYGNDRVWGKGTATASWTFDRDNAGY 662  
 Qy 214 ISAP 217  
 Db 663 AGQF 666

RESULT 2  
 US-09-306-595C-7  
 ; Sequence 7, Application US/09306595C  
 ; Patent No. 6284506  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HOSHINO, Tatsuo  
 ; APPLICANT: OJIMA, Kazuyuki  
 ; APPLICANT: SETOGUCHI, Yutaka  
 ; TITLE OF INVENTION: ISOPRENOL PRODUCTION  
 ; FILE REFERENCE: ISOPRENOL PRODUCTION  
 ; CURRENT APPLICATION NUMBER: US/09/306,595C  
 ; CURRENT FILING DATE: 1999-05-06  
 ; PRIOR APPLICATION NUMBER: 98108210  
 ; PRIOR FILING DATE: 1998-05-06  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 7  
 ; LENGTH: 1091  
 ; TYPE: PRT  
 ; ORGANISM: Phaffia rhodozyma  
 US-09-306-595C-7

Query Match 6.9%; Score 81; DB 4; Length 1091;  
 Best Local Similarity 22.4%; Pred. No. 2.5;  
 Matches 48; Conservative 30; Mismatches 50; Indels 86; Gaps 12;  
 Qy 8 WV--VEGI-----LNRLPKQFFVKCSVDWNTFVPSST-----TEKA 44  
 Db 756 WLDSVEGMEVMAASPNSYRFRLOSIRCGMAGRSYIRLATSTGDAMNMNAGKGTGA 815  
 Qy 45 ATNAMKY-----KYCWOMLVGKHQVWPVINGKKPL-----YLYGAFLMPLAK 89  
 Db 816 LETLSEYFPMOILALSGNYCIDK---KPSAINNIEGRGKSVVAESVIPAIVKSVL-- 869  
 Qy 90 ATKTT-----LNGKENLAWFTG-----GTLGLR-----KAGDWSATV 122  
 Db 870 --KTTVAIDLNLNIRKNL---IGSAMAGSIGGFNAHASDILTSIFLATQDPAQNVESM 924  
 Qy 123 RYEVYEAAL-----SVPEIDVSGIGRGNLL 146  
 Db 925 CMTLMEAVNDGKDLITCSMPAIECGTVGGGTF 958

RESULT 3  
 US-08-822-445-12  
 ; Sequence 12, Application US/08822445  
 ; Patent No. 5952223  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kaplan, Jerry  
 ; APPLICANT: Perou, Charles  
 ; APPLICANT: Moore, Karen  
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
 ; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME

NUMBER OF SEQUENCES: 32  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036/2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/822,445  
 FILING DATE: 21-MAR-1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Coruzzi, Laura A.  
 REGISTRATION NUMBER: 30,742  
 REFERENCE/DOCKET NUMBER: 7853-062-999  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-790-9090  
 TELEFAX: 212-869-8864  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3672 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 MOLECULE TYPE: protein  
 US-08-822-445-12

Query Match 6.6%; Score 77.5; DB 2; Length 3672;  
 Best Local Similarity 22.2%; Pred. No. 37;  
 Matches 41; Conservative 15; Mismatches 50; Indels 79; Gaps 7;  
 Qy 31 NTFVPSSTTTEKAATNAMKYCWOMLVGKHQVWPVINGKKPL 78  
 Db 1585 NIFLPK-----WQHLVLTLYLOQPOGKRRIHKGKISIWVSGQRKPDVT 1626  
 Qy 79 YGAFLMPLAKATKTTLNGKENLAW-FITGTLGG-----LRKAGDWSATVRYEVEALSVP 133  
 Db 1627 LDFML-----PRKTSLSDSNKTFCMIGHCLSSQEEFLQLAGKWL-----1667  
 Qy 134 EIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAFYITDLSFRAYGAYS 193  
 Db 1668 -----GNLLUFNGAKVGSQ-----EAFYLYACGNPHTSVMPCKYIG 1702  
 Qy 194 KPAND 198  
 Db 1703 KPVND 1707

RESULT 4  
 US-08-822-445-10  
 ; Sequence 10, Application US/08822445  
 ; Patent No. 5952223  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kaplan, Jerry  
 ; APPLICANT: Perou, Charles  
 ; APPLICANT: Moore, Karen  
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS  
 ; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME  
 ; NUMBER OF SEQUENCES: 32  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036/2711



```

1  RESULT      5
2  US-08-926-842B-64
3  US-08-926-842B-64
4  Sequence 64, Application US/08926842B
5  Patent No. 6030807
6  GENERAL INFORMATION:
7  APPLICANT: Sa-No. 6030807ueira, Isabel
8  APPLICANT: de Lencastre, Herminia
9  TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
10 TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
11 NUMBER OF SEQUENCES: 64
12 NUMBER OF SEQUENCES: 64
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: Klauber & Jackson
15 STREET: 411 Hackensack Avenue
16 CITY: Hackensack
17 STATE: New Jersey
18 COUNTRY: USA
19 ZIP: 07601
20 COMPUTER READABLE FORM:
21 MEDIUM TYPE: Floppy disk
22 COMPUTER: IBM PC compatible
23 OPERATING SYSTEM: PC-DOS/MS-DOS
24 SOFTWARE: PatentIn Release #1.0, Version #1.25
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: US/08/926,842B
27 FILING DATE: 10-SEP-1997
28 CLASSIFICATION: 435

```

GENERAL INFORMATION:  
APPLICANT: Knuth, Mark W.  
APPLICANT: Knoche, Kimberly K.  
APPLICANT: Selman, Susanne  
APPLICANT: Hartnett, James R.  
TITLE OF INVENTION: ISOLATED AGARASE ENZYMES FROM  
TITLE OF INVENTION: FLAVOBACTERIUM SP. STRAIN NR1  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dewitt Ross & Stevens S.C.  
STREET: 8000 Excelsior Drive, Suite 401  
CITY: Madison  
STATE: Wisconsin  
COUNTRY: U.S.A.  
ZIP: 53717-1914  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/655,704B  
FILING DATE: 03-JUN-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sara, Charles S.



REGISTRATION NUMBER: 30,492  
REFERENCE/DOCKET NUMBER: 34506.036  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-831-2100  
TELEFAX: 608-831-2106  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 367 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-655-704B-17

Query Match 6.6%; Score 76.5; DB 2; Length 367;  
Best Local Similarity 19.7%; Pred. No. 1.7; Indels 47; Gaps 7;  
Matches 37; Conservative 24; Mismatches 80; Indels 47; Gaps 7;  
QY 5 HYAWVVEGILNRLPKQFFVKCSVDWNTFVPSETSTTEKAATNAMKYKVCWQWLVGKHS 64  
Db 156 HHSFTRPTFTDYQPK-----DWSWYNDNRVTA-----NYGWDWCWNNGNRR 198  
QY 65 Q-----VPWINGCKKPLYLYGAFMLNPLAKATKTTLNGKENLAWFVGGLGG---LRKAGD 117  
Db 199 YMRMGVYVWGPKHFYYIDGQLVRVYHYNATATKYNGTWYQYF--NAMNGQFPANNANG 256  
QY 118 WSATVRYEVEALSYPEID-----VSGIGRGN-----LLKFWFAQAATAA 156  
Db 257 YTAVTYTTSSTYSPTTQAASNNNGISVIDPGNFGGAGFTKAMDIIINVEOXQWLAL 316  
QY 157 NYDPKEAN 164  
Db 317 NHTPSDAD 324

RESULT 7  
US-09-107-755-17  
; Sequence 17, Application US/09107755  
; Patent No. 6001636  
; GENERAL INFORMATION:  
; APPLICANT: Knuth, Mark W.  
; APPLICANT: Knoche, Kimberly K.  
; APPLICANT: Selman, Susanne  
; APPLICANT: Hartnett, James R.  
; TITLE OF INVENTION: ISOLATED AGARASE ENZYMES FROM  
; TITLE OF INVENTION: FLAVOBACTERIUM SP. STRAIN NR19, CLONED GENES THEREFOR,  
; TITLE OF INVENTION: AND EXPRESSION THEREOF IN TRANSFORMED HOST CELLS  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dewitt Ross & Stevens S.C.  
; STREET: 8000 Excelsior Drive, Suite 401  
; CITY: Madison  
; STATE: Wisconsin  
; COUNTRY: U.S.A.  
; ZIP: 53717-1914  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,755  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/655,704  
; FILING DATE: 03-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sara, Charles S.  
; REGISTRATION NUMBER: 30,492  
; REFERENCE/DOCKET NUMBER: 34506.036  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-831-2100  
TELEFAX: 608-831-2106  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 367 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-107-755-17

Query Match 6.6%; Score 76.5; DB 3; Length 367;  
Best Local Similarity 19.7%; Pred. No. 1.7; Indels 47; Gaps 7;  
Matches 37; Conservative 24; Mismatches 80; Indels 47; Gaps 7;  
QY 5 HYAWVVEGILNRLPKQFFVKCSVDWNTFVPSETSTTEKAATNAMKYKVCWQWLVGKHS 64  
Db 156 HHSFTRPTFTDYQPK-----DWSWYNDNRVTA-----NYGWDWCWNNGNRR 198  
QY 65 Q-----VPWINGCKKPLYLYGAFMLNPLAKATKTTLNGKENLAWFVGGLGG---LRKAGD 117  
Db 199 YMRMGVYVWGPKHFYYIDGQLVRVYHYNATATKYNGTWYQYF--NAMNGQFPANNANG 256  
QY 118 WSATVRYEVEALSYPEID-----VSGIGRGN-----LLKFWFAQAATAA 156  
Db 257 YTAVTYTTSSTYSPTTQAASNNNGISVIDPGNFGGAGFTKAMDIIINVEOXQWLAL 316  
QY 157 NYDPKEAN 164  
Db 317 NHTPSDAD 324

RESULT 8  
US-08-961-083-46  
; Sequence 46, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,083  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 241 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear



MOLECULE TYPE: protein  
US-08-961-083-46

Query Match 6.4%; Score 74.5; DB 4; Length 241;  
Best Local Similarity 24.7%; Pred. No. 1.5;  
Matches 37; Conservative 24; Mismatches 46; Indels 43; Gaps 8;  
Qy 69 INGQKKPLYL-----GAFMLNPLAKATK-TTLNGKENLAWFPGTGLGL----- 112  
Db 83 ITDERKKLYNTSPYDTASGFLVKNKSAKIKIEDLNKT-----IGVAQSI TORLITE 137  
Qy 113 --RKAGDSATVRYEYVSEALSVPEIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYK 170  
Db 138 LGKKKG-----LKRFVELGVPCL-ITSLHAHRIDTESVDRSILSGYTSKR----- 183  
Qy 171 GFSALYMYGITDSLFRAYGAYSKPANDKL 200  
Db 184 --TAL-----LDSFKPSDYGIIVTKKSNTL 207

## RESULT 9

US-08-286-819A-19  
; Sequence 19, Application US/08286819A  
; Patent No. 5871910

## GENERAL INFORMATION:

APPLICANT: ARTHUR, MICHEL  
APPLICANT: DUKTA-MALEN, SYLVIE  
APPLICANT: MOLINAS, CATHERINE  
APPLICANT: COURVALIN, PATRICE

TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR  
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/286.819A

FILING DATE: 05-AUG-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/174.682

FILING DATE: 28-DEC-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/917.146

FILING DATE: 10-AUG-1992

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR/91/00855

FILING DATE: 29-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9013579

FILING DATE: 31-OCT-1990

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5871910man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 660-060-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 19:

## SEQUENCE CHARACTERISTICS:

LENGTH: 988 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-286-819A-19

Query Match 6.3%; Score 73.5; DB 2; Length 988;

Best Local Similarity 23.0%; Pred. No. 1.6;

Matches 17; Conservative 18; Mismatches 34; Indels 5; Gaps 2;

Qy 123 RYEVVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSA--YMYGI 180

Db 234 RLEYIRGMDLETQVISHLRNLLQL---SRIGSRYEYAFRDFOENKRYSLTLYLQL 290

Qy 181 TDSLFRAYGAYSK 194

Db 291 TQELTDKAFEIHDR 304

## RESULT 10

US-08-980-357-19

; Sequence 19, Application US/08980357

; Patent No. 6013508

## GENERAL INFORMATION:

APPLICANT: ARTHUR, MICHEL

APPLICANT: DUKTA-MALEN, SYLVIE

APPLICANT: MOLINAS, CATHERINE

APPLICANT: COURVALIN, PATRICE

TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE

TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR

TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR

TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/980.357

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286.819

FILING DATE: 05-AUG-1994

APPLICATION NUMBER: US 08/174.682

FILING DATE: 28-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/917.146

FILING DATE: 10-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR/91/00855

FILING DATE: 29-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9013579

FILING DATE: 31-OCT-1990

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 6013508man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 660-060-0 PCT



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 988 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-980-357-19

Query Match 6.3%; Score 73.5; DB 3; Length 988;  
Best Local Similarity 23.0%; Pred. No. 16;  
Matches 17; Conservative 18; Mismatches 34; Indels 5; Gaps 2;

Qy 123 RYEVVEALSVPEIDVSGIGRNLKFWFAQIAAANYDPKANSFTNYKGFSA--YMYGI 180  
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :  
Db 234 RLEYIRGMDLETVOISHLHRRLLQL---SRLGSRPEYAFRDFQENKRYSLTYLLQL 290  
Qy 181 TDSLFRAYGAYSK 194  
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :  
Db 291 TQELTDKAFEIHDR 304

## RESULT 11

US-08-559-492-12  
Sequence 12, Application US/08559492  
Patent No. 5843884

GENERAL INFORMATION:  
APPLICANT: Sims, Peter J.  
TITLE OF INVENTION: C9 Complement Inhibitor  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center, 1201 West Peachtree  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/559,492  
FILING DATE: 15-NOV-1995  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: OMR154  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-873-8794  
TELEFAX: 404-873-8795

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 561 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-559-492-12

Query Match 6.2%; Score 72.5; DB 2; Length 561;  
Best Local Similarity 26.2%; Pred. No. 8.9;  
Matches 34; Conservative 20; Mismatches 53; Indels 23; Gaps 6;

Qy 84 MNPLAKA-TKTTINGKENLAWFIGTGLGRKAGDSATV-----RYEYVEA--- 129

Db 170 MDPLATPFEDNEYHGLCDRWV-DGNTLTHYRK--PWNVAVLAYETKIDKNFRTEYEQM 226  
Qy 130 ---LSVPEIDVSGIGRNLKFWFAQIAAANYD---PKEANSFTNYKGFSAIYWGTD 183  
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :  
Db 227 QAFKSIIEETSFNANLALKFTPTTEAKASKEASPKNSKLDNDKGFSSKFQFSYKN 286  
Qy 184 LSFRAYGAYS 193  
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :  
Db 287 ETYQLFLSYS 296

## RESULT 12

US-08-272-255-13  
Sequence 13, Application US/08272255  
Patent No. 5824859

GENERAL INFORMATION:  
APPLICANT: Cashmore, Anthony R.  
APPLICANT: Ahmad, Margaret  
APPLICANT: Lin, Chentao  
TITLE OF INVENTION: Blue Light Photoreceptors and Methods of  
TITLE OF INVENTION: Using the Same  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/272,255  
FILING DATE: 08-JUL-1994  
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:  
NAME: Leary Ph.D., Kathryn  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: UPN-1795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 475 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-272-255-13

Query Match 6.1%; Score 71.5; DB 2; Length 475;  
Best Local Similarity 24.3%; Pred. No. 9.1;  
Matches 36; Conservative 19; Mismatches 60; Indels 33; Gaps 7;

Qy 55 VVQMLVGHKSQVPWINGOKKPLYLGAPLMN-----PLAKATKTLNGKENL---AW 103  
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :  
Db 127 LWDQLL--HSPDQILSGSGNYSYVGPFWKNWQAQPKPTVATPTLVDSPEQLTAP 184  
Qy 104 FIGGTGLGRKAG-DWSATVRYEYVEALSVPEIDVSGIGRNLKFWFAQIAAANYDPKE 162  
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :  
Db 185 LLLSELPTLKQLGFDWDGCGFFVPEGCTAARIQE-----FCDRAIADYDPO- 231

Qy 163 ANSFTNYKGFSAIYWGTDLSFRAYG 190  
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : :  
Db 232 -RNPPAEAGTS-----GLSPALKFGAIG 253



RESULT 13

PCT-US95-08565-13  
 ; Sequence 13, Application PC/TUS9508565  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cashmore, Anthony R.  
 ; APPLICANT: Ahmad, Margaret  
 ; APPLICANT: Lin, Chentao  
 ; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris  
 ; STREET: One Liberty Place, 46th floor  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/08565  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/272,255  
 ; FILING DATE: 08-JUL-1994.  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Leary Ph.D., Kathryn  
 ; REGISTRATION NUMBER: 36,317  
 ; REFERENCE/DOCKET NUMBER: UPN-1795  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (215) 568-3100  
 ; TELEFAX: (215) 568-3439  
 ; INFORMATION FOR SEQ ID NO: 13:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 475 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PCT-US95-08565-13

Query Match 6.1%; Score 71.5; DB 5; Length 475;  
 Best Local Similarity 24.3%; Pred. No. 9.1;  
 Matches 36; Conservative 19; Mismatches 60; Indels 33; Gaps 7;  
 QY 55 VVOWLVGRKSHQVPWINGOKKPLYLGAFLLN-----PLAKATKTTLNGKENL---AW 103  
 Db 127 LWDQLL--HSPDQLSGSGNPSYVYGPFWKKNQAKPTTPVATPTELVDLSPEQLTAIAP 184  
 QY 104 FTGGTGLGRKAG-DWSATVREYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKE 162  
 Db 185 LLLSELPTLKQLGDFMDGFPVEPEGTAAIARLQE-----FCDAIADYDPO- 231  
 QY 163 ANSFNYKGFSAIYMYGTTLSLSEFRAYG 190  
 Db 232 -RNFFAEAGTS-----GLSPALKFGAIG 253

RESULT 14

US-09-540-245A-15  
 ; Sequence 15, Application US/09540245A  
 ; Patent No. 6270984  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Goodman, Corey  
 ; APPLICANT: Kid, Thomas  
 ; APPLICANT: Brose, Katja  
 ; APPLICANT: Tessier-Lavigne, Marc  
 ; TITLE OF INVENTION: Modulating Robo: Ligand Interactions

FILE REFERENCE: B98-031-3  
 ; CURRENT APPLICATION NUMBER: US/09/540,245A  
 ; CURRENT FILING DATE: 2000-03-31  
 ; PRIOR APPLICATION NUMBER: 60/065,544  
 ; PRIOR FILING DATE: 1997-11-14  
 ; PRIOR APPLICATION NUMBER: 60/081,057  
 ; PRIOR FILING DATE: 1998-04-07  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 15  
 ; LENGTH: 1395  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; US-09-540-245A-15  
 ; Query Match 6.1%; Score 71.5; DB 4; Length 1395;  
 ; Best Local Similarity 24.1%; Pred. No. 44;  
 ; Matches 26; Conservative 14; Mismatches 43; Indels 25; Gaps 5;  
 QY 68 WINGOKKPLYLGAFLLNPLAKATKTTLNGKENLAWFIGG-----TLGGLRKAGDWS 119  
 Db 569 WAKSQEKP---GA--VGPIIGYTVVEYFSPDLOTGWIVAHRVGDQVTVISGLTPTGSYV 622  
 QY 120 ATVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFT 167  
 Db 623 FLVRAENTQGISVPS-----GLSNVIK-----TIEADFDAASANDLS 659  
 RESULT 15  
 US-08-816-346-56  
 ; Sequence 56, Application US/08816346  
 ; Patent No. 6127525  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Crystal, Ronald G.  
 ; APPLICANT: Falck-Pedersen, Erik  
 ; APPLICANT: Gall, Jason  
 ; APPLICANT: Koveshdi, Imre  
 ; APPLICANT: Wickham, Thomas J.  
 ; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
 ; METHODS OF USING SAME  
 ; NUMBER OF SEQUENCES: 60  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
 ; STREET: TWO PRUDENTIAL PLAZA - 4900  
 ; CITY: CHICAGO  
 ; STATE: ILLINOIS  
 ; COUNTRY: USA  
 ; ZIP: 60601-6780  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/816,346  
 ; FILING DATE:  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; REFERENCE/DOCKET NUMBER: 67167  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/616-5600  
 ; TELEFAX: 312/616-5700  
 ; TELEX: 25-3533  
 ; INFORMATION FOR SEQ ID NO: 56:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 967 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-816-346-56



Query Match 6.0%; Score 70; DB 3; Length 967;  
Best Local Similarity 20.7%; Pred. No. 38;  
Matches 37; Conservative 25; Mismatches 61; Indels 56; Gaps 8;  
Qy 56 WQWLVGKHSQVPWINGOKKPLYLY-----GAFLMNPLAKATKTTLNKNGENLAWFIGG 107  
Db 693 WAFTRLTKETPSLGSQYDPPYTYSGSIPYLDGTFYLNHTFKKVAIT----- 739  
Qy 108 TLGGLRKAGDWSATVRYEVEALSPV---EIDVSGIGRG-----NLLKFWFAQAIANY 158  
Db 740 -----EDSSVSWFGNDRLLTPNEFEIKRSVDGEGYNVAQCNTKDWFLVQMLANY 789  
Qy 159 DPKEANSFTNYKGFSAIYMYGITDLSFRAYGAYS--KPANDKLGSDFTPRKF-DLGII 214  
Db 790 N-----IGYQGFY-----IPESYKDRWYFFRNFQPMRSRQVVDTKYKEYQQVGIL 835

Search completed: February 7, 2002, 21:36:13  
Job time: 20482 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 19:39:51 : Search time 96.2 seconds  
(without alignments)  
171.828 Million cell updates/sec

Title: US-09-391-606-7  
Perfect score: 1166  
Sequence: 1 MTKKHAYAVVVEGILNRLPKQ.....DKLGSDFTRKFDLGIISAF 217

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1162	99.7	422	2 E72042	conserved hypothet
2	1162	99.7	422	2 F86581	CHLPN 76 kda homol
3	1156	99.1	715	2 I40729	hypothetical 76k p
4	778	66.7	446	2 H71490	hypothetical prote
5	92	7.9	667	2 G69723	transketolase (EC
6	88.5	7.6	830	2 T17672	chitinase-like pro
7	84	7.2	810	2 A44054	orf1 protein - Jun
8	83	7.1	743	2 F71062	hypothetical prote
9	82.5	7.1	449	2 A54760	translational elon
10	82.5	7.1	568	2 G83558	hypothetical prote
11	82	7.0	1176	2 T18042	ice nucleation pro
12	81.5	7.0	449	2 JC5117	translation elon
13	80.5	6.9	214	2 S74663	peptide methionine
14	80.5	6.9	409	2 D70354	threonine synthase
15	80.5	6.9	469	2 T35670	hypothetical prote
16	80	6.9	644	2 H75141	oligopeptide-bind
17	79	6.8	195	2 A82416	conserved hypothet
18	79	6.8	482	2 T02355	hypothetical prote
19	79	6.8	499	2 C71895	probable outer mem
20	78.5	6.7	3796	2 T18514	lysosomal traffick
21	78	6.7	753	2 T08925	isp4 protein homol
22	77.5	6.6	250	2 S14306	chlorophyll a/b-bi
23	77.5	6.6	250	2 F64012	vacu lipoprotein h
24	77.5	6.6	476	2 T44293	glutamyl-tRNA (Gln
25	77	6.6	228	2 T39619	ubiquinol--cytochr
26	77	6.6	500	2 C69580	alpha-L-arabinofur
27	77	6.6	505	2 C69415	carbohydrate kinas
28	76.5	6.6	725	2 D81976	probable ferric si
29	76.5	6.6	755	1 LQBYPX	DNA ligase (ATP) (

ALIGNMENTS

RESULT 1  
E72042

conserved hypothetical protein CP0017 [imported] - Chlamydomydia pneumoniae (strains N; Alternate names: chlpn 76 kda homol;\_2 (ct623); hypothetical protein CPn0729 C; Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae C; Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 28-Jul-2000 C; Accession: E72042; C81623 R; Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999 A; Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis. A; Reference number: A72000; MUID:99206606 A; Accession: E72042 A; Molecule type: DNA A; Residues: 1-422 <ARN> A; Cross-references: GB:AE001654; GB:AE001363; NID:94377031; PIDN:AA018868.1; PID:943 A; Experimental source: strain CML029 R; Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hick, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salz Nucleic Acids Res. 28, 1397-1406, 2000 A; Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR3 A; Reference number: A81500; MUID:20150255 A; Accession: C81623 A; Status: preliminary A; Molecule type: DNA A; Residues: 1-422 <REA> A; Cross-references: GB:AE002165; GB:AE002161; NID:97188948; PIDN:AAF37913.1; PID:971 A; Experimental source: strain AR39, HL cells C; Comment: This sequence was originally identified as homologous to part of a sequen PIR:H71490). C; Genetix: A; Gene: CPn0729; CP0017 C; Superfamily: Chlamydia trachomatis hypothetical protein CT623

Query Match 99.7%; Score 1162; DB 2; Length 422;  
Best Local Similarity 99.5%; Pred. No. 3.6e-97;  
Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MTKKHYAVVVEGILNRLPKQFVKCSVVDWNTFVPSSTTEKAATNAMYKYCVQWMLV	60
Db	206	MTKKHYAVVVEGILNRLPKQFVKCSVVDWNTFVPSSTTEKAATNAMYKYCVQWMLV	265
Qy	61	GKHSQVPWINGOKKPLYLYGAFNLPLAKATKTTLNGENLAWFTGGTLGGLRKAGDWSA	120
Db	266	GKHSQVPWINGOKKPLYLYGAFNLPLAKATKTTLNGENLAWFTGGTLGGLRKAGDWSA	325
Qy	121	TVRYEYVVALSVPEIDVSGIGRGNLLKFWFAQIAAANTDPKEANSFTNYKGFSAIYMYGI	180
Db	326	TVRYEYVVALSVPEIDVSGIGRGNLLKFWFAQIAAANTDPKEANSFTNYKGFSAIYMYGI	385
Qy	181	TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF	217

hypothetical prote  
hypothetical prote  
ligand gated chann  
probable glutamate  
probable ligand-ga  
probable ubiquitin  
hypothetical prote  
TonB-dependent rec  
ribonucleotide red  
hypothetical prote  
carboxylesterase (  
mannosyl-glycoprot  
threonine synthase  
hypothetical prote  
gene 112 protein -



Db 386 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 422

RESULT 2

CHLPN 76 kba homolog\_2 (CT623) [imported] - Chlamydia pneumoniae (strain J138)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C:Accession: F86581

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349

A:Accession: F86581

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-422 <STO>

A:Cross-references: GB:BA000008; NID:g8979101; PIDN:BAA98936.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

C:Superfamily: Chlamydia trachomatis hypothetical protein CT623

Query Match 99.7%; Score 1162; DB 2; Length 422;

Best Local Similarity 99.5%; Pred. No. 3.6e-97;

Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMYKVCWQWLV 60

Db 206 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMYKVCWQWLV 265

Qy 61 GKHSQVPWINGOKKPLYLYGAFLLNPLAKATKTTLNKENLAWFTGGTLGLRKAGDWSA 120

Db 266 GKHSQVPWINGOKKPLYLYGAFLLNPLAKATKTTLNKENLAWFTGGTLGLRKAGDWSA 325

Qy 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYMYGI 180

Db 326 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYMYGI 385

Qy 181 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217

Db 386 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 422

RESULT 3

I40729

hypothetical 76k protein - Chlamydia pneumoniae (strain AR39)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 31-Mar-2000

C:Accession: I40729

R:Perez-Melgosa, M.; Kuo, C.

Infect. Immun. 62, 880-886, 1994

A:Title: Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76-kil

A:Reference number: I40729; MUID:94156481

A:Accession: I40729

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-715 <RES>

A:Cross-references: GB:L23921; NID:g435961; PIDN:AAA23117.1; PID:g435962

A:Experimental source: strain AR-39

C:Comment: This is the hypothetical translation of a sequence that was reported as two s

Query Match 99.1%; Score 1156; DB 2; Length 715;

Best Local Similarity 99.5%; Pred. No. 2.4e-96;

Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMYKVCWQWLV 60

Db 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMYKVCWQWLV 60

Qy 61 GKHSQVPWINGOKKPLYLYGAFLLNPLAKATKTTLNKENLAWFTGGTLGLRKAGDWSA 120

Db 61 GKHSQVPWINGOKKPLYLYGAFLLNPLAKATKTTLNKENLAWFTGGTLGLRKAGDWSA 120

Qy 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYMYGI 180

Db 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYMYGI 180

Qy 181 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 216

Db 181 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 216

RESULT 4

H71490

hypothetical protein CT623 - Chlamydia trachomatis (serotype D, strain UW3/Cx)

N:Alternate names: chlpn 76kda homolog CT623

C:Species: Chlamydia trachomatis

C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 28-Jul-2000

C:Accession: H71490

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitch

Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia

A:Reference number: A71570; MUID:99000809

A:Accession: H71490

A:Molecule type: DNA

A:Residues: 1-446 <ARN>

A:Cross-references: GB:AE001333; GB:AE001273; NID:g3329068; PIDN:AAC68227.1; PID:g33

A:Experimental source: serotype D, strain UW-3/Cx

C:Comment: This sequence was originally identified as homologous to part of a sequen

PIR:E72042).

C:Genetics:

A:Gene: CT623

C:Superfamily: Chlamydia trachomatis hypothetical protein CT623

Query Match 66.7%; Score 778; DB 2; Length 446;

Best Local Similarity 63.8%; Pred. No. 1.6e-62;

Matches 148; Conservative 26; Mismatches 42; Indels 16; Gaps 4;

Qy 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFV-----PSETSTTEKAATNAMYKVCV 55

Db 216 MAEREYAWVVEAILNKLPGNFVVTSTVVDWNTLAKTNDPADASAAQPAKPT-KYDYLV 274

Qy 56 WQWLVGKHSQVPWINGOKKPLYLYGAFLLNPLAK-----ATKTT--LNGKENLAWFI 105

Db 275 WQWLVGKSTAMPFNGQTKNLYTGAVLFNPLAEIPENWKQSTTPTTKITNGKENHAWFI 334

Qy 106 GTTGLGLRKAGDWSATVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANS 165

Db 335 GCSLGGVRRAGDWSATVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANS 394

Qy 166 FTNYKGFSAALYMYGITDLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217

Db 395 FTNYKGVSYQFVGMGLTDSVSFRAYAAKSPANDNLGSDFTYRKYDGLGISF 446

RESULT 5

G69723

transketolase (EC 2.2.1.1) tkt - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000

C:Accession: G69723; S57401

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.,

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,

Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scan

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S



```

Query Match      7.6%; Score 88.5; DB 2; Length 830;
Best Local Similarity 22.4%; Pred. No. 5;
Matches 38; Conservative 19; Mismatches 44; Indels 69; Gaps 6;

QY 4 KHVAWYVEGLNR-----LPKQFFVKCSVVDNTEFVPSSETSTT-----EKAATNAMKY 51
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 50 ENTTFEGEDLVKRGKVTWPK-----DNNMSIPACTWIIIPFGVKYALPGNLJKY 100

```

```

RESULT      8
F71062
hypothetical protein PH1196 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: F71062
R: Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida,
DNA Res. 5, 55-75, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-ther-
A:Reference number: A71000; MUID:98344137
A:Accession: F71062
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-743 <RAW>
A:Cross-references: GB:A0000005; NID:g3236132; PIDN:BAA30296.1; PID:g3257613
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced
C:Genetics:
A:Gene: PH1196
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1196

```

Query Match	7.1%	Score 83;	DB 2;	Length 743;
Best Local Similarity	18.5%;	Pred. No. 14;		
Matches 42;	Conservative	35;	Mismatches	76;
Indels	74;			Gaps 9;



RESULT 10  
G83558  
hypothetical protein PA0696 [imported] - Pseudomonas aeruginosa (strain PA01)

```

Query Match      7.0%; Score 82; DB 2: Length 1176;
Best Local Similarity 23.6%; Pred. No. 29;
Matches 43; Conservative 26; Mismatches 61; Indels 53; Gaps
Qy 59 LVGRHSQVPWINGCQKKPLLYLGAFLLMNLAKAKTTTLNGKENLAWTIG----- 106
Db 179 IIGHAGENLLINGNE--LVLTIGAY-----ACQNTNGSRDV--FVGKACQNAQLSD 227
Qy 107 ---GTLGGLR-----KAGDWSATVRYEYEAALSVPEDIVSGIGRGNLLKFWF 150

```



Db 228 SVAIGRCAGLNTISGNNVWVGSGSPAATAK-----ETVAIGA--KAGYNAGNSLQSVF 281  
Qy 151 AQAIAANYDPKREANSFTNYKGFSAIYMGITDSLSFRAYGAYSPANDKLGSDTERKFD 210  
Db 282 V-GRSAGYNGGTSTFGIYEA-----GYGATDSLSLYNTFVGY-KTGNIRTCSFNTIMCAD 334  
Qy 211 LG 212  
Db 335 AG 336

## RESULT 12

JC5117  
translation elongation factor eEF-1 alpha - Trypanosoma cruzi  
C:Species: Trypanosoma cruzi  
C:Date: 02-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 02-Feb-2001  
C:Accession: JC5117  
R:Billaut-Mulot, O.; Fernandez-Gomez, R.; Loyens, M.; Ouaisi, A.  
Gene 174, 19-26, 1996  
A:Title: Trypanosoma cruzi elongation factor 1-alpha: Nuclear localization in parasites  
A:Reference number: JC5117; MUID:97017123  
A:Accession: JC5117  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-449 <BIL>  
A:Cross-references: GB:U76077; NID:gi929444; PIDN:AAC01751.1; PID:gi929445  
C:Comment: This protein plays roles in genetic control of cell death and nuclear process  
C:Genetics:  
A:Gene: EF-1alpha  
C:Superfamily: translation elongation factor Tu; translation elongation factor Tu homolog  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:8-156/Domain: translation elongation factor Tu homology <ETU>  
F:14-21,91-94,153-156/Region: GTP binding #status predicted  
F:14-21/Region: nucleotide-binding motif A (P-loop)  
F:162-181/Region: actin binding #status predicted  
F:230-266/Domain: tRNA binding #status predicted <TRN>

Query Match 7.0%; Score 81.5; DB 2; Length 449;  
Best Local Similarity 21.6%; Pred. No. 10;  
Matches 42; Conservative 25; Mismatches 72; Indels 55; Gaps 9;  
Qy 19 KOFFVKCSVDP-----WNTFVPSETSTTEKAATNAMKYKCV---NQW--LVCKH 63  
Db 146 KQMVCCNMDDKSVNFAQRIYDEIVKVSAYLKKRGVNVKVRFPISPGWGNMIDKS 205  
Qy 64 SQVPRINGQKPLYLGAFLNPLAKATKTTINGKENLAWFIGGTGTLG-----GLRK 114  
Db 206 ENMPYKG---PTLLEALDMLPEPPVRPSDKPLRLPLQDVYKIGG-IGTVPVGRVETGTMK 261  
Qy 115 AGD-----WSATVRYEVE-----ALSVPEIDVSGIGRGNLLKFWFAQA 153  
Db 262 PGDVVTFAPAVNTTEKVSIMHHEQLAEATPGDNVGNVKNVSKVRDIRRGNVC-----G 315  
Qy 154 IAAAYDPKREANSFT 167  
Db 316 NSKNDPPKREADFT 329

## RESULT 13

S74663  
peptide methionine sulfoxide reductase msrA - Synechocystis sp. (strain PCC 6803)  
N:Alternate names: hypothetical protein slr1795  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S74663  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.  
A:Reference number: S74322; MUID:97061201

A:Accession: S74663  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-214 <KAN>  
A:Cross-references: EMBL:D90900; GB:AB001339; NID:gi1651768; PIDN:BAAL6815.1; PID:gi16  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Gene: msrA  
A:Start codon: GTG  
C:Superfamily: peptide methionine sulfoxide reductase

Query Match 6.9%; Score 80.5; DB 2; Length 214;  
Best Local Similarity 24.6%; Pred. No. 5.1;  
Matches 29; Conservative 16; Mismatches 34; Indels 39; Gaps 6;  
Qy 80 GAFLMNLAKA-----TKTTLNCKENLAWFIGGTGGLRK----- 114  
Db 8 GSFLISPSKVIPIPVVDINPVSTTARTEK-AVFAGGCFWGLEAMFEFVRGVKDVQGY 66  
Qy 115 AGDWSATVRY-----EYVEALSVPEDVSGIGRGNLLKFWFAQAIAAAYDPKEAN 164  
Db 67 SGGTEANTANYARVSGGTDHAESEIEI-VYDPAQVSYGELLAKIFF-----SVGHDPTQVN 119

## RESULT 14

D70354  
threonine synthase (EC 4.2.99.2) thrC1 [similarity] - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 26-May-2000  
C:Accession: D70354  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666  
A:Accession: D70354  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-409 <AQF>  
A:Cross-references: GB:AE000698; NID:gi2983224; PIDN:AAC06822.1; PID:gi2983225; GB:AE0  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: thrC1  
C:Superfamily: threonine dehydratase  
C:Keywords: carbon-oxygen lyase; phosphoprotein; pyridoxal phosphate  
F:109/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 6.9%; Score 80.5; DB 2; Length 409;  
Best Local Similarity 23.2%; Pred. No. 11;  
Matches 33; Conservative 20; Mismatches 54; Indels 35; Gaps 6;  
Qy 78 LYGAFLMNLAKATKTTINGKENLAWFIGGTGTLG-----GLRK-AGDWSATVRYEVEAL 130  
Db 164 IIGSLVFNPTVVAVEGNYDDVNRCLSEIANDLGWAFVNIINIRPYPAESKTLAEVAQL 223  
Qy 131 S--VPEIDVSGIGRGNL-LKFWFAQAIAAAYDPKEANSFTNYKGFSAIYMGITDSLSFR 187  
Db 224 GWRAPDVVVVAPAAAGSLYTKIW-----KGFNELKKVGLIDEVKTR 263  
Qy 188 AYGAYSKRPAND-----KLGSDP 204  
Db 264 MYGAQAGCCSPIAQNAREGRDF 285

## RESULT 15

T35670  
hypothetical protein SC7B7.02 SC7B7.02 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35670  
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.



submitted to the EMBL Data Library, November 1997

A:Reference number: Z21586

A:Accession: F35670

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-469 <MUR>

A:Cross-references: EMBL:AL009199; PIDN:CAAL5782.1; GSPDB:GN000070; SCOEDB:SC7B7.02

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC7B7.02

```

Query Match          6.9%; Score 80.5; DB 2; Length 469;
Best Local Similarity 24.2%; Pred. No.13;
Matches 45; Conservative 20; Mismatches 68; Indels 53; Gaps 6;

Qy 10 VEGILNRLPKOFFVKCSVVDHNTFVPSETSTTEKAAATNAMKYKVCWOWLVGKHQVPI 69
Db 260 VGVVVK-----VNETVVKQMFVPGSGTQFQKA-----QAIS 293

Qy 70 NGQKKPLYLYGAFLMNPLAKATKT-----TLNGKENLAWFIGTGLGRLKAGDWS 119
Db 294 NDQKALLYPSGGWIENEMKATKADFOQTGIPSMITLTDKPALPY-----EALRAAG-- 345

Qy 120 ATVRVEYVEALSVPEIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAIYMG 179
Db 346 -----EPFIVPKQGNPAGGKREVLRLAMLSKAAANFS-KTKLAPTIVKGTVPADGYG 396

Qy 180 ITDSLS 185
Db 397 STALVS 402

```

Search completed: February 7, 2002, 21:38:03  
Job time: 7092 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:32:46 ; Search time 76.51 seconds  
(without alignments)  
103.990 Million cell updates/sec

Title: US-09-391-606-7

Perfect score: 1166

Sequence: 1 MTKKHAYWVEGILNRLPKQ.....DKLGSDFTRFKDLGIISAF 217

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	7.9	667	TKT_BACSU	P45694 bacillus su
2	84	7.2	810	COAT_JCDNV	Q90053 junonia coe
3	81.5	7.0	449	EF1A_TRYBB	P41166 trypanosoma
4	80	6.9	225	MTRO_METKA	Q32864 methanopyru
5	80	6.9	1786	1 YCF1_ARATH	P56785 arabidopsis
6	77.5	6.6	250	1 VACJ_HAEIN	P44042 haemophilus
7	77.5	6.6	476	1 COAT_BACSD	Q929X0 bacillus ha
8	77.5	6.6	3801	1 LYST_HUMAN	Q99698 homo sapien
9	77	6.6	228	1 UCRI_SCHPO	Q09154 schizosacch
10	77	6.6	500	1 ABFA_BACSU	P94531 bacillus su
11	77	6.6	1109	1 CSAA_BACTF	Q9X682 bacillus th
12	76.5	6.6	755	1 DNLI_YEAST	P04819 saccharomyc
13	76	6.5	811	1 COAT_GMDNV	Q90125 galleria me
14	75.5	6.5	431	1 NORE_CHLPN	Q92723 chlamydia p
15	75	6.4	542	1 ESTS_DROVI	Q05487 drosophila
16	75	6.4	789	1 LF1A_DROME	P11995 drosophila
17	75	6.4	1311	1 STRH_STRPN	P49610 streptococc
18	74.5	6.4	825	1 EGLS_HANAN	P06835 hansenuia a
19	74	6.3	700	1 NONA_DROME	Q04047 drosophila
20	74	6.3	789	1 LP1B_DROME	P11996 drosophila
21	73.5	6.3	315	1 OMPP_ECOLI	P34210 escherichia
22	73.5	6.3	445	1 DS74_CAEEL	P45971 caenorhabdi
23	73.5	6.3	456	1 ONAL_CHLTR	Q84252 chlamydia t
24	73.5	6.3	752	1 PSAA_ODOSI	P49479 odontella s
25	73.5	6.3	988	1 TNP6_ENTFC	Q06238 enterococcu
26	73	6.3	530	1 YA9A_SCHPO	Q09788 schizosacch
27	73	6.3	591	1 VFYA_SCHPO	Q9ut18 schizosacch
28	72.5	6.2	283	1 SOHA_RAT	P22789 rattus norv
29	72.5	6.2	291	1 THIG_CYACA	Q19915 cyanidium c
30	72.5	6.2	557	1 CO9_RABIT	P48747 erythrolagus
31	72.5	6.2	572	1 PTLB_STAAN	P11162 staphylococ
32	72.5	6.2	809	1 COAT_DSDNV	Q71155 diatraea sa
33	72.5	6.2	931	1 TRP6_HUMAN	Q9y210 homo sapien

## RESULT 1

TKT_BACSU	TKT_BACSU	STANDARD;	PRT;	667 AA.
ID	AC	P45694;		
DT	01-NOV-1995	(Rel. 32, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	20-AUG-2001	(Rel. 40, Last annotation update)		
DE	TRANSEKTOLOASE (EC 2.2.1.1).			
GN	TKT OR TKTA.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RA	Rose M., Entian K.;			
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 561-667 FROM N.A.			
RC	STRAIN=168;			
RA	Schiott T., von Wachenfeldt C., Hederstedt L.;			
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE + D-XYULOSE 5-PHOSPHATE.			
CC	3-PHOSPHATE - D-RIBOSE 5-PHOSPHATE + D-XYULOSE 5-PHOSPHATE.			
CC	-1- COFACTOR: THIAMINE PYROPHOSPHATE.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
DR	EMBL; Z73234; CAA97616.1;			
DR	EMBL; X87845; CAA61113.1;			
DR	EMBL; Z99113; CAB13673.1;			
DR	HSP; P23254; IAYO.			
DR	Subtilist; BG11247; tkt.			
DR	InterPro; IPR000360; Transketolase.			
DR	Pfam; PF00455; transketolase.1.			
DR	PROSITE; PS00801; TRANSKETOLASE_1; 1.			
DR	PROSITE; PS00802; TRANSKETOLASE_2; 1.			
KW	Transferase; Thiamine pyrophosphate; Complete proteome.			
SQ	SEQUENCE 667 AA; 72344 MW; D93BCAQD246148AF CRC64;			

Query Match 7.9%; Score 92; DB 1; Length 667;

Best Local Similarity 22.9%; Pred. No. 0.57;

Matches 50; Conservative 32; Mismatches 68; Indels 68; Gaps 12;

QY 1 MTKKHAYWVEGILNRLPKQFFVKCSVVVDWNTFVPSFTSTTEKAATNAMYKY-----C 54

DB 273 LTRKAYAWTYE-----EDFIVFSEVYEHFAVAVKESGKKEQWNAQFAKYEVYPPEL 325



QY 55 VQWLVGKHQSVWINGOKPLLYLVGAFMLNPLAKATKTTLNG--KENLAWFIGGT--LGG 111  
 DB 326 AEQLELAIKGELPKDWDQEVYVEKGGSLAS--RASSEVLNGLAKKIPFVVGGSADLAG 383  
 QY 112 -----LRKAGWSATVRYEYVEALSVPEDVSGIGRGNLKFV-----FAQAIANYDPK 161  
 DB 384 SNKTIKNAGDFTA-----VDYSG-----KNWFVGVRREFAMGAALN---- 419  
 QY 162 EANSFTNYKFGS---ALYMYGITSLSFRAYGAYSKPA 196  
 DB 420 -----GMLHGLGVFGGT-----FFVFSYDLRPA 444

## RESULT 2

ID COAT\_JCDNV STANDARD; PRT; 810 AA.  
 AC Q90053;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE COAT PROTEIN VP1 (STRUCTURAL PROTEIN VP1) (CONTAINS: COAT PROTEIN VP  
 DE (STRUCTURAL PROTEIN VP2); COAT PROTEIN VP3 (STRUCTURAL PROTEIN VP3);  
 DE COAT PROTEIN VP4 (STRUCTURAL PROTEIN VP4)).  
 GN VP.  
 OS Junonia coenia densovirus (JCDNV).  
 OC Viruses; ssDNA viruses; Parvoviridae; Densovirinae; Densovirus.  
 OX NCBI\_TaxID=12524;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=93033112; PubMed=14113502;  
 RX Dumas B., Jourdan M., Pascaud A.M., Bergoin M.;  
 RT "Complete nucleotide sequence of the cloned infectious genome of  
 RT Junonia coenia densovirus reveals an organization unique among  
 RT parvoviruses";  
 RL Virology 191:202-222(1992).  
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
 CC OF A COMBINATION OF VP4, VP3, VP2 AND VP1 (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: THE FOUR DIFFERENT COAT PROTEINS ARE  
 CC PRODUCED BY ALTERNATIVE INITIATION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; S47266; AAB23698.1; -  
 DR InterPro; IPR003433; denso\_VP4.  
 DR Pfam; PF02336; denso\_VP4; 1.  
 KW Coat protein; Alternative initiation.  
 FT CHAIN 1 810 COAT PROTEIN VP1.  
 FT CHAIN 278 810 COAT PROTEIN VP2.  
 FT CHAIN 323 810 COAT PROTEIN VP3.  
 FT CHAIN 374 810 COAT PROTEIN VP4.  
 SQ SEQUENCE 810 AA; 87895 MW; 0D618F04ADD04DE4 CRC64;

Query Match 7.2%; Score 84; DB 1; Length 810;  
 Best Local Similarity 25.8%; Pred. No. 3.9;  
 Matches 61; Conservative 22; Mismatches 59; Indels 94; Gaps 18;

QY 2 TKKHVAVWVGELNRLPKQFFVKCSVDWNTFVPSSTTEKAATNAMKYKVCWQMLVG 61  
 DB 413 TKSH-KFMIFGLAN-----NVIGTIGTGT---AVNRI-ITLCL----- 446  
 QY 62 KHSQVPWINGOKPLLYLYCA-----FLMNPLA-----KATK-TTLN 96  
 DB 447 --AEIPL--OKLPLMYNQSEFDLLPGSRVVECNKVKVIFRTNRIAPETSTATKQATLN 501  
 QY 97 KENLAWFIGTGLGLRAGDWSATVRYEYVEALSVPEDVSGIGRGNLKFWEFAQAI-- 154

DB 502 QISNLQ-----TAVGLNKLK-W-----GIDR-SFTAFQSDQPMIP 534  
 QY 155 -----AANYDKPEANSFTNYKGFSAHYMYGITSLSFRAYGAYSKPANDKIGSDPTF 206  
 DB 535 TATSAPKYEY--ITGTTGYRGMIAHY-YG-ADSTNDAAFGNAGNYPHQVGS-FTF 585

## RESULT 3

ID EFIA\_TRYBB STANDARD; PRT; 449 AA.  
 AC P41166;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).  
 GN TEFL.  
 OS Trypanosoma brucei brucei.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LVH/75/USAMRU-K/18;  
 RX MEDLINE=94364999; PubMed=8083206;  
 RA Kaur K.J., Ruben L.;  
 RT "Protein translation elongation factor-1 alpha from Trypanosoma  
 RT brucei binds calmodulin";  
 RL J. Biol. Chem. 269:23045-23050(1994).  
 RN [2]  
 RP SEQUENCE OF 18-415 FROM N.A.  
 RA MEDLINE=94089672; PubMed=8265589;  
 RX Baldauf S.L., Palmer J.D.;  
 RT "Animals and fungi are each other's closest relatives: congruent  
 RT evidence from multiple proteins";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:11558-11562(1993).  
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF  
 CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN  
 CC BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-TU/EF-1A SUBFAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; L25868; AAA16602.1; -  
 DR EMBL; U10562; AAA57476.1; -  
 DR HSP: Q01698; ITUI  
 DR InterPro; IPR000795; GTP\_EFTU.  
 DR Pfam; PF00009; GTP\_EFTU; 1.  
 DR PRINTS; PR00315; ELONGATNFCT.  
 DR PROSITE; PS00301; EFATOR\_GTP; 1.  
 KW Elongation factor; Protein biosynthesis; GTP-binding.  
 FT NP\_BIND 14 21 GTP (BY SIMILARITY).  
 FT NP\_BIND 91 95 GTP (BY SIMILARITY).  
 FT NP\_BIND 153 156 GTP (BY SIMILARITY).  
 FT CONFLICT 240 242 TCT -> DVT (IN REF. 2).  
 SQ SEQUENCE 449 AA; 49033 MW; F21113FB80A5116A CRC64;

Query Match 7.0%; Score 81.5; DB 1; Length 449;  
 Best Local Similarity 21.5%; Pred. No. 3.5;  
 Matches 42; Conservative 25; Mismatches 71; Indels 57; Gaps 9;

QY 19 KQFVVKCSVD-----WNTFVPSSTTEKAATNAMKYKVC---WQW--LVGKH 63  
 DB 146 QMWVCCNKMKDDKTVNGQERYDEIVKVSAYIKKVGYNVEKRVFVPSIGQGNMIEKS 205



Qy 64 SOVPWINGQKPLYLYGAFLMP-----LAKATKTTINGKENLAWFIGTGLGLR 113  
 Db 206 EKMPYKG---PTLLEALDLEPPVRPSDKPLRLPLOTCTKIGGIGTVP--VGRVETGVM 260  
 Qy 114 KAGD-----NSATRYEYVE-----ALSVPEIDVSGIGRGNLLKFWFAQ 152  
 Db 261 RPDGVVTFAPANVTTEVKSIEHHEQLAEATPGDGVNKFVNRKNSVDIRRGVNC----- 314  
 Qy 153 ALAANYDPKEANSFT 167  
 Db 315 GNTKNDPPKEADFT 329

## RESULT 4

MTRD\_METKA STANDARD; PRT; 225 AA.  
 AC O32864;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE TETRAHYDROMETHANOPTERIN S-METHYLTRANSFERASE SUBUNIT D (EC 2.1.1.86)  
 DE (N5-METHYL-TETRAHYDROMETHANOPTERIN--COENZYME M METHYLTRANSFERASE  
 DE SUBUNIT D).  
 DN MTRD.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyrales; Methanopyrus.  
 OX NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98121200; PubMed=9461302;  
 RA Hays U., Thauer R.K.;  
 RT Identification of the active site histidine in the corrinoid protein  
 RT MTR of the energy-conserving methyltransferase complex from  
 RT Methanobacterium thermoautotrophicum.";  
 RL Eur. J. Biochem. 250:783-788(1997).  
 CC -!- FUNCTION: THIS ENZYME COMPLEX CATALYZES AN INTERMEDIATE STEP IN  
 CC METHANOGENESIS, THE FORMATION OF METHYL-COENZYME M AND  
 CC TETRAHYDROMETHANOPTERIN FROM COENZYME M AND N5-METHYL-  
 CC TETRAHYDROMETHANOPTERIN. THIS IS A ENERGY-CONSERVING, SODIUM-ION  
 CC TRANSLOCATING STEP.  
 CC -!- CATALYTIC ACTIVITY: 5-METHYL-5,6,7,8-TETRAHYDROMETHANOPTERIN +  
 CC 2-MERCAPTOETHANESULFONATE = 5,6,7,8-TETRAHYDROMETHANOPTERIN +  
 CC 2-(METHYLTHIO)ETHANESULFONATE.  
 CC -!- SUBUNIT: COMPOSED OF 8 DIFFERENT SUBUNITS (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Y14428; CAA74768.1; -  
 KW Transferase; Methyltransferase; Transmembrane; Methanogenesis.  
 FT TRANSMEM 5 25 POTENTIAL.  
 FT TRANSMEM 40 60 POTENTIAL.  
 FT TRANSMEM 67 87 POTENTIAL.  
 FT TRANSMEM 132 152 POTENTIAL.  
 FT TRANSMEM 161 181 POTENTIAL.  
 FT TRANSMEM 204 224 POTENTIAL.  
 SQ SEQUENCE 225 AA; 22872 MW; 6FA9546897670D36 CRC64;

Query Match 6.9%; Score 80; DB 1; Length 225;  
 Best Local Similarity 23.9%; Pred. No. 2.3;  
 Matches 44; Conservative 22; Mismatches 72; Indels 46; Gaps 8;

Qy 38 TSTTEKAATN-----AMKY---YCVQWLVGKHSQVPWINGQKPLYLYGAFLMN 85  
 Db 42 TQTTQLAAGSLTGLTAAANSQKPLVLNWLNGALGAATMTMLVGNFIYVYGCGPP 101

Qy 86 PLAKATKTTINGKENLAW-----FIGGTGLGLRKAGDSATRYEYVEALS 131  
 Db 102 CSAVKDKDPITGWDOEAYVTPGTECHGIPTVSFVSGILGGL-LGSGGAMVYIYALKVLG 160  
 Qy 132 VPEIDVSGIGRGNLLKFWFAQIAANY-----DPKEANSFTNY-KGFSALYMYGI 180  
 Db 161 MSAALAGLWANG-----FFYANAVLASYNIGGTIEGYHDPK-----FTRLPLKAVVCSLVEGI 212  
 Qy 181 TDSL 184  
 Db 213 VASV 216

## RESULT 5

YCF1\_ARATH STANDARD; PRT; 1786 AA.  
 AC P56785;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL 213.7 KDA PROTEIN YCF1.  
 DN YCF1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20039611; PubMed=10574454;  
 RA Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;  
 RT "Complete structure of the chloroplast genome of Arabidopsis  
 RT thaliana";  
 RL DNA Res. 6:283-290(1999).  
 CC -!- FUNCTION: NOT YET KNOWN.  
 CC -!- SIMILARITY: BELONGS TO THE YCF1 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AP000423; BA844445.1; -  
 DR EMBL; AP000423; BA844433.1; -  
 KW Chloroplast; Hypothetical protein.  
 SQ SEQUENCE 1786 AA; 213727 MW; CFFD2A4D776D7E5D CRC64;

Query Match 6.9%; Score 80; DB 1; Length 1786;  
 Best Local Similarity 19.9%; Pred. No. 21;  
 Matches 33; Conservative 37; Mismatches 74; Indels 22; Gaps 6;

Qy 8 WVVEGILNRLPKQFVKCVVDW---NTEVPSETSTTEKAATNAMKYKC---VWQWLVGK 62  
 Db 840 WLTGDIQIKILFPFYLK----PWHKSKFQASQKARLKTGDKGKNDFCFLTIV---GM 891  
 Qy 63 HSQVPWINGQKPLYLYGAFLMNPLAKATKTTINGKENLAWFIGTGLGLRKAGDSATV 122  
 Db 892 ETELPFGSAQRKPSF-----FEPISKELKKRIKKKKKSFV---LKFIERAPIFLKV 942  
 Qy 123 RYEYVEALSVPEIDVSGIGRGNLLKFWFAQIAANYDPKEANSFTN 168  
 Db 943 AKETKNWLKNFIFIKGISKRNLIPLFGPREIYELNEPKKDSITSN 988

## RESULT 6

VACJ\_HAEIN STANDARD; PRT; 250 AA.  
 ID VACJ\_HAEIN



```
AC P44042;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACJ LIPOPROTEIN HOMOLOG PRECURSOR.
GN VACJ OR HI0718.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae RD.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32754; AAC22375.1;
DR TIGR; HI0718;
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Signal; Complete proteome.
FT SIGNAL 1 18
FT CHAIN 19 250 VACJ LIPOPROTEIN HOMOLOG.
FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 250 AA; 28110 MW; 7DDEC4FC587091BB CRC64;

Query Match 6.68; Score 77.5; DB 1; Length 250;
Best Local Similarity 29.68; Pred. No. 4.3;
Matches 32; Conservative 10; Mismatches 35; Indels 31; Gaps 6;

QY 28 VDMNFTV--PSETSTTEKAATNAM---KYKCVWQWLVGKHSQVFWINGQKKPLYLYGAF 82
DB 144 VDAGTIVYVLPYNATTPQLTCAVDVDAAYMYPFQWVG-----PW-----ALVKYGVQ 192

QY 83 LNNPLAKATKTLNKENLAWFIGTGLGRLKAGDWSATVRYVEAL 130
DB 193 AVDARAK-----NLNNAEL-----LRQAQPYITFREAYQNL 225

RESULT 7
ID GATB_BACHD STANDARD; PRT; 476 AA.
AC Q949X0; Q93PV7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-) (GLU-ADT
DE SUBUNIT B).
GN GATB OR BH0667.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
```

```
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=99184646; PubMed=10086842;
RA Takami H., Nakasone K., Ogasawara N., Hirama C., Nakamura Y.,
RA Masui N., Fuji F., Takaki Y., Inoue A., Horikoshi K.;
RT "Sequencing of three lambda clones from the genome of alkaliphilic
RT Bacillus sp. strain C-125.";
FT Extremophiles 3:29-34(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC GLN-TRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-
CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP
CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GATB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011836; BAA75312.1;
DR EMBL; AP001509; BAB04386.1;
DR InterPro; IPR001773; Gln_amidotransf_B.
DR Pfam; PF01162; PET112; 1.
DR PROSITE; PS01234; GATB; 1.
KW Protein biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 476 AA; 53414 MW; C212B522413C579B CRC64;

Query Match 6.68; Score 77.5; DB 1; Length 476;
Best Local Similarity 27.28; Pred. No. 8.6;
Matches 25; Conservative 13; Mismatches 39; Indels 15; Gaps 2;

QY 90 ATKTTL-----NGKENLAWFIGTGLGRLKAGDWSATVRYE-----YVEALSYPE 134
DB 253 ANKTVLMRVKSGSDDYRFPEPDLVALHIDDEWKARISEIPELDPARKKRYVEELGLPA 312

QY 135 IDVSGIGRGNLLKFWFAQIAANYDPKEANSF 166
DB 313 YDAMVLTITLKEMSDFFETIAKAGADPKLASNW 344

RESULT 8
ID LYST_HUMAN STANDARD; PRT; 3801 AA.
AC Q99698; Q99709; O43274;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LYSOSOMAL TRAFFICKING REGULATOR (BEIGE HOMOLOG).
GN CHS1 OR LYST OR CHS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```



RN CC  
 RX MEDLINE-97051925; PubMed-8896560;  
 RA Nagle D.L., Karim M.A., Woolf E.A., Holmgren L., Bork P., Misumi D.J.,  
 RA McGrail S.H., Dussault B.J., Perou C.M., Bolissy R.E., Duyk G.M.,  
 RA Spritz R.A., Moore K.J.;  
 RT "Identification and mutation analysis of the complete gene for  
 RT Chediak-Higashi syndrome.";  
 RL Nat. Genet. 14:307-311(1996).  
 RN [2]  
 RN CC  
 RX MEDLINE-96353977; PubMed-8717042;  
 RA Barbosa M.D.F.S., Nguyen Q.A., Tchernev V.T., Ashley J.A.,  
 RA Better J.C., Blaydes S.M., Brandt S.J., Chotai D., Hodgman C.,  
 RA Solari R.C.E.S., Lovett M., Kingsmore S.F.;  
 RT "Identification of the homologous beige and Chediak-Higashi syndrome  
 RT genes.";  
 RL Nature 382:262-265(1996).  
 RN [3]  
 RN CC  
 RX MEDLINE-97358584; PubMed-9215680;  
 RA Barbosa M.D.F.S., Barrat F.J., Tchernev V.T., Nguyen Q.A.,  
 RA Mishra V.S., Colnan S.D., Pastural E., Dufourcq-Lagelouse R.,  
 RA Fischer A., Holcombe R.F., Wallace M.R., Brandt S.J.,  
 RA de Saint Basile G., Kingsmore S.F.;  
 RT "Identification of mutations in two major mRNA isoforms of the  
 RT Chediak-Higashi syndrome gene in human and mouse.";  
 RL Hum. Mol. Genet. 5:1091-1098(1997).  
 RN [4]  
 RN CC  
 RX MEDLINE-97358584; PubMed-9215680;  
 RA Barbosa M.D.F.S., Barrat F.J., Tchernev V.T., Nguyen Q.A.,  
 RA Kingsmore S.F., de Saint Basile G.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: MAY REGULATE LYSOSOME TRANSPORT.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2  
 CC AND 3; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN ADULT AND FETAL  
 CC THYMUS, PERIPHERAL BLOOD LEUKOCYTES, BONE MARROW AND SEVERAL  
 CC REGIONS OF THE ADULT BRAIN.  
 CC -!- DISEASE: DEFECTS IN CHS1 ARE THE CAUSE OF CHEDIAK-HIGASHI SYNDROME  
 CC (CHS) A RARE AUTOSOMAL RECESSIVE DISORDER CHARACTERIZED BY  
 CC HYPOPIGMENTATION, SEVERE IMMUNOLOGIC DEFICIENCY, A BLEEDING  
 CC TENDENCY, NEUROLOGIC ABNORMALITIES, ABNORMAL INTRACELLULAR  
 CC TRANSPORT TO AND FROM THE LYSOSOME, AND GIANT INCLUSION BODIES IN  
 CC A VARIETY OF CELL TYPES.  
 CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
 CC -!- SIMILARITY: CONTAINS 1 BEACH DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U84744; AAB87737.1; .  
 DR EMBL; U67615; AAB41309.1; .  
 DR EMBL; U72192; AAB39697.1; .  
 DR EMBL; L77889; AAB51608.1; .  
 DR EMBL; U70064; AAB41533.1; .  
 DR MIM; 214500; .  
 DR InterPro; IPR000409; Beige\_BEACH.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF02138; Beach; 1.  
 DR Pfam; PF00400; WD40; 5.  
 DR ProDom; PD007848; Beige\_BEACH; 1.  
 DR SMART; SM00320; WD40; 4.  
 DR PROSITE; PS50197; BEACH; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 1.

DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Protein transport; Transport; Repeat; WD repeat; Alternative splicing.  
 FT DOMAIN 31 36 POLY-GLU.  
 FT DOMAIN 2448 2452 POLY-LEU.  
 FT REPEAT 562 700 WD 1.  
 FT REPEAT 1582 1626 WD 2.  
 FT DOMAIN 3139 3422 BEACH.  
 FT REPEAT 3563 3602 WD 3.  
 FT REPEAT 3614 3653 WD 4.  
 FT REPEAT 3656 3699 WD 5.  
 FT REPEAT 3700 3744 WD 6.  
 FT REPEAT 3749 3788 WD 7.  
 FT VARSPLIC 1515 1531 ESDRPEGAEYINPGERL -> GMMTGLSDLYTKIVFRL  
 FT VARSPLIC 1532 3801 MISSING (IN ISOFORM 3).  
 FT VARSPLIC 1988 2001 VCRSFVKIIAEVLG -> MARSFRKKGQSCT (IN  
 FT VARSPLIC 2002 3801 MISSING (IN ISOFORM 2).  
 SQ SEQUENCE 3801 AA; 429117 MW; 288B370AD4BDAF11 CRC64;  
 Query Match 6.6%; Score 77.5; DB 1; Length 3801;  
 Best Local Similarity 22.2%; Pred. No. 82;  
 Matches 41; Conservative 15; Mismatches 50; Indels 79; Gaps 7;  
 QY 31 NTEVPSETSTTEKAATNAMYKYCYVMQVLVGRKHSQV-----WINGQKKPLYL 78  
 Db 1585 NIEFSPK-----WQHLVLYLQOPQCKRRIHKGKISIWVSQRRKPDVT 1626  
 QY 79 YGAFLMPLAKATKTLNGKENLAW-FIGTTLGG-----LRKAGDWSATVRYEVEALSVP 133  
 Db 1627 LDFML-----PRKTSLSDSNKTFCMIGHCLSSOEFLQLAGKWDL----- 1667  
 QY 134 EIDVSGICRGNLLKFEWQAIAAAYNDPREANSFTNYKGFSAALYMYGIDTSLSFAYGAYS 193  
 Db 1668 -----GNLLFNGAKVSGQ-----EAYLYACGPNHTSVMPCKYKG 1702  
 QY 194 KPAND 198  
 Db 1703 KPVND 1707  
 RESULT 9  
 UCRL\_SCHPO  
 ID UCRL\_SCHPO STANDARD; PRT: 228 AA.  
 AC Q09154; O42942;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE UBIOQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT, MITOCHONDRIAL  
 DE PRECURSOR (EC 1.10.2.2) (RIESKE IRON-SULFUR PROTEIN) (RISP).  
 GN RIPI OR SPBC16H5.06.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96279043; PubMed-8663290;  
 RA Dirago J., Bruel C., Graham L.A., Slonimski P., Trumpower B.L.;  
 RT "Heterologous complementation of a Rieske iron-sulfur protein-  
 RT deficient Saccharomyces cerevisiae by the ripl gene of  
 RT Schizosaccharomyces pombe.";  
 RL J. Biol. Chem. 271:15341-15345(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RL Wood V., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: COMPONENT OF THE UBIOQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX) WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL



```

CC COUPLED TO ATP SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q +
CC 2 FERROCYTOCHROME C.
CC -1- SUBUNIT: FUNGI BCL COMPLEX CONTAINS 10 SUBUNITS; 3 RESPIRATORY
CC SUBUNITS, 2 CORE PROTEINS AND 5 LOW-MOLECULAR WEIGHT PROTEINS.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC -1- MISCELLANEOUS: THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S
CC PROTEIN.
CC -1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
CC BACTERIAL, CHLOROPLAST).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U40480; AAC49359.1; -
CC EMBL; AL022104; CAA17904.1; -
CC HSSP; P13272; 1RIE.
CC InterPro; IPR001700; RNA_pol_A_bac.
CC Pfam; PF00355; Rieseke; 1.
CC PRINTS; PR00162; RIESKE.
CC ProDom; PD001179; RNA_pol_A_bac; 1.
CC PROSITE; PS00199; RIESKE_1; 1.
CC PROSITE; PS00200; RIESKE_2; 1.
CC Mitochondrion; Electron transport; Respiratory chain; Iron-sulfur;
CC Oxidoreductase; Inner membrane; Transmembrane; Transit peptide.
CC TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
CC CHAIN ? 228 UBIQUINOL-CYTOCHROME C REDUCTASE IRON-
CC SULFUR SUBUNIT.
CC METAL 172 172 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC METAL 174 174 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC METAL 191 191 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC METAL 194 194 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC FT DISULFID 177 193 BY SIMILARITY.
CC FT CONFLICT 183 183 A -> R (IN REF. 1).
CC SEQUENCE 228 AA; 24740 MW; C50CC8BA159E31B2 CRC64;

Query Match 6.6%; Score 77; DB 1; Length 228;
Best Local Similarity 30.0%; Pred. No. 4.3;
Matches 27; Conservative 17; Mismatches 28; Indels 18; Gaps 4;

QY 91 TKTTLNGENLAWFTGTLGLGRKAG-----DSSATVRYEYVEALSVPEDVSGI 140
Db 58 TKSTGDRSRVTSYAVGTMGALTAAGAAQATVHDFLASWSASAD---VLAMSKAEVDLSKI 114
QY 141 GRG-NLLKFWFAQAIAANY----DPKEANS 165
Db 115 PGKKNLVKWKQKPFVIRHPTPEEQEANS 144

RESULT 10
ID ABFA_BACSU STANDARD; PRT; 500 AA.
AC P94531; Q05096;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ALPHA-L-ARABINOFURANOSIDASE 1 (EC 3.2.1.55) (ARABINOSIDASE).
GN ABFA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124191; PubMed=8969504;

```

```

RA Wibat A., Carter N., Brignell C.S., Guy J.B., Piper K.,
RA Sanders J., Emmerson P.T., Harwood C.R.;
RT "The dnaB-phcA (256 degrees-240 degrees) region of the Bacillus
RT subtilis chromosome containing genes responsible for stress
RT responses, the utilization of plant cell walls and primary
RT metabolism.";
RL Microbiology 142:3067-3078(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97237725; PubMed=9084180;
RA Sa-Nogueira I., Nogueira T.V., Soares S., de Lencastre H.;
RT "The Bacillus subtilis L-arabinose (ara) operon: nucleotide sequence,
RT genetic organization and expression.";
RL Microbiology 143:957-969(1997).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING ALPHA-L-
CC ARABINOFURANOSIDE RESIDUES IN ALPHA-L-ARABINOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 51 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z75208; CAA9595.1; -
CC EMBL; X89810; CAA61937.1; -
CC EMBL; Z99118; CAB14832.1; -
CC Subtilist; BG11900; abfa.
KW Hydrolase; Glycosidase; Complete proteome.
FT CONFLICT 114 114 P -> A (IN REF. 2).
SQ SEQUENCE 500 AA; 57060 MW; 7397ED52A4987686 CRC64;

Query Match 6.6%; Score 77; DB 1; Length 500;
Best Local Similarity 23.1%; Pred. No. 10;
Matches 50; Conservative 33; Mismatches 89; Indels 44; Gaps 13;

QY 15 NLRPKQFFVKCSVDWNTFVPSSTTE--KAATNAMK-YKYCVWMLVGRKHSQ----- 65
Db 250 NNLPNYL---ARSMDLDFIKSVAATCDYVRAKTRSKRTINLSLDENWVYHSNEADKKV 306
QY 66 VPWLNGOK--KPLY-----LYGAFLMNPLAKATKTLNGKENLAWFICGTLGLGRKAGD 117
Db 307 EPWITAREILEDYVNFEDALLVGSLLITMLQHADRKVKTAQLAVNVNVIAPIM--TEKGE 364
QY 118 -WSATVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPK-EANSFTNYKGFSA 175
Db 365 AWRQPIFYPMHA-----SVYGRGSLKPLIS-----SPKYDCSDFTDVPYVDA 409
QY 176 YMWG-ITDSLSFRAYGAYSKPANDKLGSDFTFRKFD 210
Db 410 VVYSEEBETLTIFAV---NKAEDQMETEISLRGFE 441

RESULT 11
ID CSAA_BACTF STANDARD; PRT; 1109 AA.
AC Q9X682;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PESTICIDIAL CRYSTAL PROTEIN CRY28AA (INSECTICIDAL DELTA-ENDOTOXIN
DE DE CRYXXVIIIA(A)) (CRYSTALLINE ENTOMOCIDAL PROTOXIN) (126 KDA CRYSTAL
DE PROTEIN).
GN CRY28AA OR CRYXXVIIIA(A).
OS Bacillus thuringiensis (subsp. finitimus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=29337;
RN [1]

```



```
RP SEQUENCE FROM N.A.
RC STRAIN-VKPM B-1161;
RX MEDLINE=99330166; PubMed=10403372;
RA Wojciechowska J.A., Lewitin E., Revina L.P., Zalunin I.A.,
RA Chestukhina G.G.;
RT "Two novel delta-endotoxin gene families cry26 and cry28 from Bacillus
RL thuringiensis ssp. finitimus.";
RL FEBS Lett. 453:46-48(1999).
CC !- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC EPITHELIAL CELLS OF INSECTS.
CC !- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPOROULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC !- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC !- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF132928; AAD24189.1; -.
DR InterPro; IPR001178; Endotoxin.
DR Pfam; PF00555; endotoxin; 1.
KW Toxin; Sporulation.
SQ SEQUENCE 1109 AA; 125712 MW; 10C80705508F5CDA CRC64;

Query Match      6.6%; Score 77; DB 1; Length 1109;
Best Local Similarity 27.0%; Pred. No. 24;
Matches 51; Conservative 23; Mismatches 67; Indels 48; Gaps 12;

QY 15 NRPLOKQFFVKCSVDNWFVPSST--TEKAATNAMYKKYCV--WQMLVGK-----H 63
DB 387 NEYKFNFNVN-----NONEPQETTPNDYGGNSQKFKHLSHFLLIHKLEFAEYFH 440

QY 64 S--QVPPW---INGOKKPLYLGYFLMPLAKATKTLNGKENLAWFTGGTLGGLR---- 113
DB 441 SIFALGWTHNSVNSN--LISESVSTQIPLVKAYEVTNNSVIRGPGFTGGDLIEURDKCS 498

QY 114 ---KAGD---WSATRYEYVEALSPEIDVSGIGRGNLLKFWFAQAIANYDPKEANSFT 167
DB 499 IKCKASSLLKYYAISLFYAANNAIAV-SIDVGDSGAGVLLQPTFSR-----KGNNTFT 549

QY 168 ----NYKGF 172
DB 550 IODLNYKDF 558

RESULT 12
DNLI_YEAST
ID DNLI_YEAST STANDARD; PRT; 755 AA.
AC P04819; Q12736;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA LIGASE I, MITOCHONDRIAL PRECURSOR (EC 6.5.1.1)
DE (POLYDEOXYRIBONUCLEOTIDE SYNTHASE [ATP]).
GN CDC9 OR YDL164C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86093646; PubMed=3909103;
RA Barker D.G., White J.H.M., Johnston L.H.;
RT "The nucleotide sequence of the DNA ligase gene (CDC9) from
RT Saccharomyces cerevisiae: a gene which is cell-cycle regulated and
```

```
RT induced in response to DNA damage.";
RL Nucleic Acids Res. 13:8323-8337(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Pohl T.M.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 610-755 FROM N.A.
RX STRAIN=AH22;
RX MEDLINE=93247548; PubMed=8483449;
RA Wehner E.P., Rao E.P., Brendel M.;
RT "Molecular structure and genetic regulation of SFA, a gene
RT responsible for resistance to formaldehyde in Saccharomyces
RT cerevisiae, and characterization of its protein product.";
RL Mol. Gen. Genet. 237:351-358(1993).
RN [4]
RP ALTERNATIVE INITIATION.
RX MEDLINE=20003237; PubMed=10531002;
RA Miller M., Rainey M., Pullen T., Stirling C.J.;
RT "The yeast CDC9 gene encodes both a nuclear and a mitochondrial form
RT of DNA ligase I.";
RL Curr. Biol. 9:1085-1094(1999).
CC !- FUNCTION: THIS PROTEIN SEALS NICKS IN DOUBLE-STRANDED DNA DURING
CC DNA REPLICATION, DNA RECOMBINATION AND DNA REPAIR. THE
CC MITOCHONDRIAL FORM IS REQUIRED FOR MITOCHONDRIAL DNA MAINTENANCE
CC BUT IS NON-ESSENTIAL WHILE THE NUCLEAR FORM IS ESSENTIAL FOR CELL
CC VIABILITY.
CC !- CATALYTIC ACTIVITY: ATP + (DEOXYRIBONUCLEOTIDE)(N) +
CC (DEOXYRIBONUCLEOTIDE)(M) = AMP + PYROPHOSPHATE +
CC (- SUBCELLULAR LOCATION: NUCLEAR AND MITOCHONDRIAL.
CC !- ALTERNATIVE PRODUCTS: A SINGLE GENE PRODUCES BOTH A MITOCHONDRIAL
CC FORM AND A NUCLEAR FORM BY THE USE OF ALTERNATIVE INITIATION
CC CODONS IN THE SAME READING FRAME.
CC !- MISCELLANEOUS: CDC9 IS INCLUDED WITHIN THE CATEGORY OF SO-CALLED
CC 'START GENES', ENCODING PROTEINS WHICH ARE REQUIRED IN EARLY G1,
CC WHEN THE CELL IS FACED WITH THE OPTION OF INITIATING A FURTHER
CC CELL CYCLE.
CC !- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03246; CAA27005.1; -.
DR EMBL; 267750; CAA91582.1; -.
DR EMBL; 274212; CAA98737.1; -.
DR EMBL; X68020; CAA48158.1; -.
DR PIR; S05830; LQYPPX.
DR PIR; S31138; S31138.
DR SGD; S0002323; CDC9.
DR InterPro; IPR000977; DNA_ligase.
DR Pfam; PF01068; DNA_ligase; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR PROSITE; PS00333; DNA_LIGASE_A2; 1.
DR PROSITE; PS0160; DNA_LIGASE_A3; 1.
KW DNA repair; DNA replication; DNA recombination; Cell division; Ligase;
KW ATP-binding; Nuclear protein; Mitochondrion; Alternative initiation;
KW Transit peptide.
FT TRANSIT 1 44 MITOCHONDRION (POTENTIAL).
FT CHAIN 45 755 DNA LIGASE I, MITOCHONDRIAL ISOFORM.
FT CHAIN 24 755 DNA LIGASE I, NUCLEAR ISOFORM.
FT INIT_MET 24 24 FOR NUCLEAR ISOFORM.
FT BINDING 419 419 AMP (BY SIMILARITY).
FT CONFLICT 69 69 D -> E (IN REF. 1).
FT CONFLICT 186 186 L -> V (IN REF. 1).
FT CONFLICT 671 671 G -> E (IN REF. 3).
FT CONFLICT 724 724 R -> I (IN REF. 3).
```



```
SQ SEQUENCE 755 AA; 84828 MW; B7C2ECAF5C61CAE7 CRC64;
Query Match 6.6%; Score 76.5; DB 1; Length 755;
Best Local Similarity 21.08; Pred. No. 18; Indels 53; Gaps 7;
Matches 45; Conservative 26; Mismatches 90;
QY 2 TKKHAYVVEGILNRLP-----KQFFVKSVVDW-----TFVPSETSTTEK 43
DB 434 TMRYSRNGENWTERYPENITDFIQDLDTTKNLILDCFAVWDKQDKILPFOVLSTRK 493
QY 44 ---AATNAMKYKCVQWMLVKHKSQVPWINGOKKPLYLYGAFLMNPPLAKATKTTLNGKEN 100
DB 494 RKDVELNDVKVKVCLFAFDILCYNDERLINKSKERREY-----LTKVKVVPGEFQY 546
QY 101 LAWFITGTLGLRKAGDSATVRYE--YVEALSVPEDVSGIGRGNLLKFWFAQAIARA 158
DB 547 AFQITTNLDELQKFLDESVNHSCEGLMKMLEGPE-----SHY 585
QY 159 DPKEANSFTNYKGFSAIYMYGITSLSFRAYGAY 192
DB 586 EPSKRS--RNWLKLLKDYLEGVDSLDLCVLGAY 617
RESULT 13
COAT_GMDNV STANDARD; PRT; 811 AA.
AC 090125; 090128; 090126; 090127;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE COAT PROTEIN VP1 (STRUCTURAL PROTEIN VP1) [CONTAINS: COAT PROTEIN VP
DE (STRUCTURAL PROTEIN VP2); COAT PROTEIN VP3 (STRUCTURAL PROTEIN VP3);
DE COAT PROTEIN VP4 (STRUCTURAL PROTEIN VP4)].
GN VP1.
OS Galleria mellonella densovirus (GMDNV).
OC Viruses; ssDNA viruses; Parvoviridae; Densovirinae; Densovirus.
OX NCBI_TaxID=37138;
RN [1]
RP SEQUENCE FROM N.A.
RA Tijssen P.;
RT "Organization and expression of the ambisense genome of
RT densoleucosis virus of Galleria mellonella (GMDNV).";
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (3.6 ANGSTROMS) OF 396-811.
RX MEDLINE=99036856; PubMed=9817847;
RA Simpson A.A., Chipman P.R., Baker T.S., Tijssen P., Rossmann M.G.;
RT "The structure of an insect parvovirus (Galleria mellonella
RT densovirus) at 3.7 A resolution.";
RL Structure 6:1355-1367(1998).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC OF A COMBINATION OF VP4, VP3, VP2 AND VP1.
CC -!- ALTERNATIVE PRODUCTS: THE FOUR DIFFERENT COAT PROTEINS ARE
CC PRODUCED BY ALTERNATIVE INITIATION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L32896; AAA66966.1; -
CC DR EMBL; L32896; AAA66964.1; -
CC DR EMBL; L32896; AAA66965.1; -
CC DR EMBL; L32896; AAA66967.1; -
CC DR PDB; 1DNV; 16-FEB-99.
CC DR InterPro; IPR003433; denso_VP4.
CC DR Pfam; PF02336; denso_VP4; 1.
KW Coat protein; Alternative initiation; 3D-structure.
FT CHAIN 1 811 COAT PROTEIN VP1.
```

```
FT CHAIN 277 811 COAT PROTEIN VP2.
FT CHAIN 324 811 COAT PROTEIN VP3.
FT CHAIN 375 811 COAT PROTEIN VP4.
SQ SEQUENCE 811 AA; 88542 MW; 0A61B09172DF99B5 CRC64;
Query Match 6.5%; Score 76; DB 1; Length 811;
Best Local Similarity 22.4%; Pred. No. 21;
Matches 59; Conservative 36; Mismatches 82; Indels 86; Gaps 16;
QY 2 TKKHAYVVEGILNRLPKQFFVKSVVDWNTFFPSETSTTEKATNAMKYKCVQWMLVG 61
DB 414 TKSH-KFMIFLAN-----NVIGPTGTGT---AVNRL-----LTT 445
QY 62 KHSQVPWINGOKKPLYLYGA-----FLMNPPLAKAT-----KTTLN 96
DB 446 CLAEIPW---QKPLLYNQSEFDLLPFSRVVVECNVVFETNRIATSETSTVTKQATLN 502
QY 97 GKENLAWFIQGTGLRKAGDSATVRYEVE-----ALSVPEID-VSG-IG-RGNL 145
DB 503 QISNVQ-----TAIQLNKLKLG-WGINRAFTAFQSDQPMIPTATTAPKYEPTGDTGYRGM 556
QY 146 LKFWFAQAI-----AANYDPKEANSFTNYKGFSAIYMY-----GITDLSLFRAYGAYS 193
DB 557 ADYTGADSTNDTAFGNAGNYPHGVSSFTFLQNYCYMYQQTNGTGWPCLAHLQOQFDS 616
QY 194 KPANDKLGSDFTFRKFDLGLIISA 216
DB 617 KTVNQCLIDVY-KPKMGLIKS 638
RESULT 14
ID_NORF_CHLPN STANDARD; PRT; 431 AA.
AC 097723; 097955;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE NA(+)-TRANSLLOCATING NADH-QUINONE REDUCTASE SUBUNIT F
DE (EC 1.6.5.-) (NA(+)-TRANSLLOCATING NOR SUBUNIT F) (NA(+)-NOR SUBUNIT F)
DE (NOR COMPLEX SUBUNIT F) (NOR-1 SUBUNIT F).
GN NORF OR NOR6 OR CPN0883 OR CP0983.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RN [3]
RP Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
```



Nucleic Acids Res. 28:2311-2314(2000).

1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. THE FIRST STEP IS CATALYZED BY NQR, WHICH ACCEPTS ELECTRONS FROM NADH AND REDUCES UBIQUINONE-1 TO UBISEMIQUINONE BY A ONE-ELECTRON TRANSFER PATHWAY (BY SIMILARITY).

2- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) + UBIQUINOL + NA(+) (OUT).

3- COFACTOR: FAD AND A 2FE-2S CLUSTER (BY SIMILARITY).

4- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE AND NQRF (BY SIMILARITY).

5- SUBCELLULAR LOCATION: INNER MEMBRANE (POTENTIAL).

6- SIMILARITY: BELONGS TO THE NQR FAMILY.

7- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

8- EMBL: AE001669; AAD19021.1; -

9- EMBL: AE002236; AAF38762.1; -

10- EMBL: AP002548; BAA99091.1; -

11- HSSP: P00235; 1FRR.

12- TIGR: CP0983; -

13- InterPro: IPR001041; Ferredoxin.

14- InterPro: IPR001433; Oxidored\_FAD.

15- Pfam: PF00111; fer2; 1.

16- Pfam: PF00175; NAD\_binding; 1.

17- PROSITE: PS00197; 2FE2S\_FERREDOXIN; FALSE\_NEG.

18- Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;

19- Flavoprotein; FAD; Iron-sulfur; Transmembrane; Inner membrane;

20- Complete proteome.

21- TRANSMEM 7 29 POTENTIAL.

22- DOMAIN 45 123 FERREDOXIN.

23- DOMAIN 289 413 CATALYTIC.

24- METAL 76 76 IRON-SULFUR (2FE-2S) (BY SIMILARITY).

25- METAL 82 82 IRON-SULFUR (2FE-2S) (BY SIMILARITY).

26- METAL 85 85 IRON-SULFUR (2FE-2S) (BY SIMILARITY).

27- METAL 117 117 IRON-SULFUR (2FE-2S) (BY SIMILARITY).

28- CONFLICT 4 4 L -> P (IN REF. 3).

29- SEQUENCE 431 AA; 48500 MW; D6D9A7A474C62708 CRC64;

Query Match 6.5%; Score 75.5; DB 1; Length 431;

Best Local Similarity 26.0%; Pred. No. 12;

Matches 57; Conservative 33; Mismatches 84; Indels 45; Gaps 15;

Qy 34 VPS-ETSTTEKAATNAMKYCYVQWL-----VGKHSQVPWINGOKK-PLYLYGAFLMNP 86

Db 178 VPSYKTNSSDWKQTAPEY-YSDWEHFLDQVIDNSQLPADSANKAYSLASYPAL--P 234

Qy 87 LAK-----ATKTLNGKEN--LAWFIGTGLGLRAGDWSATRVVEYVEAL-----SVPEI 135

Db 235 TIKFNIRIATPPFINGKPSIEMGVCSSYVFLSKPGD-KITVSGPYGESEFMKDDRLPI 293

Qy 136 DV-----SGIGRGNL-----KWF-AQAIAN-YDPKEANSFTNYKGSALY 176

Db 294 FLIGAGSFGSRSHLDLLLNKHSKREIDLWYGARSLKENIYQEYENLERQFPNFH--Y 351

Qy 177 MYGITDLSFRAYGAYSKPANDKLGSDFTFRKFDLGIIS 215

Db 352 HLVLSEPLPDIAAGWDK--DDPTKTNFLFRAFLNQLS 388

RESULT 15

ESTS\_DROVI

ID ESTS\_DROVI

AC Q05487;

DT 01-JUN-1994 (Rel. 29, Created)

01-JUN-1994 (Rel. 29, Last sequence update)

01-FEB-1996 (Rel. 33, Last annotation update)

ESTERASE S PRECURSOR (EC 3.1.1.1) (EST-S) (CARBOXYLIC-ESTER HYDROLASE).

OS Drosophila virilis (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCHI\_TaxID=7244;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93347990; PubMed=8346032;

RA Sergeev P.V., Yenikolopov G.N., Peunova N.I., Kuzin B.A., Khechumian R.A., Korochkin L.I., Georgiev G.P.;

RT "Regulation of tissue-specific expression of the esterase S gene in Drosophila virilis.";

RL Nucleic Acids Res. 21:3545-3551(1993).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=89392127; PubMed=2783177;

RA Sergeev P.V., Castillo J.E., Peunova N.I., Yenikolopov G.N.;

RT "Primary structure of the esterase S gene from Drosophila virilis.";

RL Bioorg. Khim. 15:839-843(1989).

CC 1- FUNCTION: TRANSFERRED FROM THE EJACULATORY BULBS OF MALES TO THE FEMALE GENITALS UPON COPULATION, PLAYS AN IMPORTANT ROLE IN THE REPRODUCTIVE BIOLOGY.

CC 1- CATALYTIC ACTIVITY: A CARBOXYLIC ESTER + H(2)O = AN ALCOHOL + A CARBOXYLIC ANION.

CC 1- SUBUNIT: MONOMER.

CC 1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN THE EJACULATORY BULBS OF MALE.

CC 1- DEVELOPMENTAL STAGE: IN THE MALE, IT APPEARS 3 DAYS AFTER EMERGENCE IN THE IMAGO STAGE AND REACHES MAXIMUM LEVELS BY THE 10TH DAY.

CC 1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X70351; CAA49809.1; -

DR PIR: JN0438; JN0438.

DR PIR: S32019; S32019.

DR PIR: S34853; S34853.

DR HSSP: P21836; LMAH.

DR FlyBase: FBgn0013077; Dvir\Ests.

DR InterPro: IPR002018; Carboxylesterase\_B.

DR Pfam: PF00135; Coesterase; 1.

DR PROSITE: PS00122; CARBOXYLESTERASE\_B\_1; 1.

DR PROSITE: PS00941; CARBOXYLESTERASE\_B\_2; 1.

KW Hydrolase; Serine esterase; Glycoprotein; Signal.

FT SIGNAL 1 22

FT CHAIN 23 542

FT ACT\_SITE 204 204

FT DISULFID 81 100

FT DISULFID 256 268

FT DISULFID 507 528

FT CARBOHYD 110 110

FT CARBOHYD 396 396

FT SEQUENCE 542 AA; 61069 MW; 9134648A7B573F84 CRC64;

Query Match 6.4%; Score 75; DB 1; Length 542;

Best Local Similarity 26.0%; Pred. No. 17;

Matches 46; Conservative 23; Mismatches 46; Indels 62; Gaps 13;

Qy 61 GKHSQVPWINGOKKPLYLYGAFLMNPPLAKATKTTNGRE-----NLAW-----FIG--G 107



Db 320 GNFAQVPLASYTPENGIYNAALL--LAK-----DANGKERIEELNTRWNELAPYFLAYPY 373  
QY 108 TLGGLRKAGDSA---TVRYEYVEALSVPEIDVSGIGRCGNLLKFWFAQAIAANYDPKEAN 164  
Db 374 TL-----KRSEMNHAHSOKLYQYL-----GYKNFSVNVYFDVQRL----- 408  
QY 165 SFTNYKGFSAIYMWGITDSL-SFRAYGA-----YSKPANDKLGSDFTFRKFDLGI 213  
Db 409 -FTN-----ELYKKGIELSLDSHRKHGASPVYAYVDNPA-DKSLAQFLAKRSDISL 458

Search completed: February 7, 2002, 21:42:39  
Job time: 593 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: February 7, 2002, 21:34:34 ; Search time 137.02 Seconds  
(without alignments)  
264.895 Million cell updates/sec

Title: US-09-391-606-8  
Perfect score: 2412  
Sequence: 1 MVNPIGPGIDERTPPAD.....QKLSEEDLNSAVDHHHHH 490

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2392	99.2	490	21 AAY71957	C. pneumoniae 76 k
2	2188	90.7	651	21 AAY71954	Chlamydia pneumoni
3	2184	90.5	452	21 AAY71956	3'-truncated Chlam
4	2162	89.6	478	20 AAY35358	Chlamydia pneumoni
5	1850	76.7	583	21 AAY71955	5'-truncated Chlam
6	281.5	11.7	350	20 AAY37571	Chlamydia trachoma
7	231.5	9.6	361	21 AAB13695	Chlamydia sp. prot
8	231.5	9.6	361	22 AAG83263	Protein encoded by
9	182	7.5	1822	13 AAR27745	Extracellular fact
10	164.5	6.8	331	20 AAY37572	Chlamydia trachoma
11	153.5	6.4	1237	21 AAY81609	Streptococcus pneu

12	153	6.3	600	22 AAG91443	C glutamicum prote
13	152	6.3	1561	17 AAW02098	S. mutans antigen
14	150.5	6.2	643	16 AAR84568	Trypanosoma cruzi
15	148	6.1	571	21 AAG29572	Arabidopsis thalia
16	148	6.1	729	22 AAB19849	Mycobacterium tube
17	148	6.1	881	22 AAG70752	S cerevisiae apopt
18	147.5	6.1	1095	22 AAG83030	S. epidermidis ope
19	147	6.1	1566	16 AAR79643	Immunodominant ant
20	144	6.0	5024	22 AAG82935	S. epidermidis ope
21	142.5	5.9	1279	22 AAG83047	S. epidermidis ope
22	142	5.9	1607	19 AAW50897	Mouse laminin G1 c
23	141.5	5.9	605	20 AAY21870	Amino acid sequenc
24	141.5	5.9	654	20 AAY21859	Amino acid sequenc
25	140.5	5.8	564	16 AAR84565	Trypanosoma cruzi
26	140	5.8	876	22 AAG91035	C glutamicum prote
27	138	5.7	666	20 AAY21871	Amino acid sequenc
28	138	5.7	715	20 AAY21860	Amino acid sequenc
29	138	5.7	1576	21 AAB19802	Human laminin 2 ma
30	138	5.7	1576	21 AAB48453	Human laminin 2 ga
31	138	5.7	1584	21 AAB19804	Human laminin 2 ga
32	138	5.7	1609	19 AAW50898	Human laminin G1 c
33	138	5.7	1609	21 AAB19801	Human laminin 2 ga
34	138	5.7	1609	21 AAB48452	Human laminin 8 po
35	138	5.7	1617	21 AAB19803	Human laminin 2 ga
36	138	5.7	2742	21 AAB23012	Human APC protein
37	138	5.7	2842	15 AAG63508	Adenomatous polyo
38	138	5.7	2843	13 AAR26052	APC gene product 1
39	138	5.7	2843	15 AAR58634	Adenomatous polyo
40	138	5.7	2843	16 AAW11922	Adenomatous polyo
41	138	5.7	2843	18 AAW35392	Human adenomatous
42	138	5.7	2843	19 AAW76140	Human APC protein
43	138	5.7	2843	19 AAW76144	Human APC protein
44	138	5.7	2843	19 AAW38370	Human adenomatous
45	138	5.7	2843	21 AAB23011	Human APC protein

## ALIGNMENTS

RESULT: 1  
AAY71957  
ID AAY71957 standard; Protein; 490 AA.  
XX AC AAY71957;  
XX DT 26-MAR-2001 (first entry)  
XX DE C. pneumoniae 76 kDa protein truncation mutant fusion protein.  
XX KW 76 kDa protein; bactericidal; diagnosis; prevention;  
KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;  
KW acute respiratory disease; cough; sore throat; hoarseness; fever;  
KW vaccine; immunisation; treatment; fusion protein; truncation mutant;  
KW muten.  
XX OS Chlamydia pneumoniae.  
OS Synthetic.  
XX FH Key  
XX FT Region  
XX FT Location/Qualifiers  
FT 453..490  
FT /note= "This part of the sequence is unrelated  
FT to the C. pneumoniae 76 kDa protein"

XX WO2000066739-A2.  
XX PD 09-NOV-2000.  
XX PF 03-MAY-2000; 2000WO-CA00511.  
XX PR 03-MAY-1999; 99US-0132270.  
XX PR 30-JUN-1999; 99US-0141276.  
XX PA (AVET ) AVENTIS PASTEUR LTD.



```

XX PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX DE
XX KW WPI; 2000-687542/67.
XX DR N-PSDB; AAD02066.
XX PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX PT useful for vaccinating against Chlamydia infections -
XX OS Claim 33; Fig 3; 112pp; English.
XX PN
XX CC The present sequence is a fusion protein comprising a
XX CC truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded
XX CC residues. C. pneumoniae 76 kDa protein is used in the
XX CC diagnosis, prevention and treatment of C. pneumoniae
XX CC infections (e.g. pneumonia, upper respiratory
XX CC tract disease, bronchitis, sinusitis and acute respiratory
XX CC disease such as cough, sore throat, hoarseness, fever; and
XX CC abnormal chest sounds on auscultation). C. pneumoniae sequence
XX CC is also used as vaccines for immunising humans against diseases
XX CC caused by C. pneumoniae.
XX CC
XX SQ Sequence 490 AA;

Query Match 99.2%; Score 2392; DB 21; Length 490;
Best Local Similarity 99.2%; Pred. No. 5.8e-153;
Matches 486; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MWNPIGPGPIDETERTPPADLSAAGLEASAAKSAEQAQRIAGAAKPKESKTSVERWSI 60
DB 1 mwnpiggpiddetertppadlsaggleasaanksaeeqriagaekpkesktdsverwsi 60
QY 61 LRSVAVNLMSTADKLGITASSNSSTSRSDVDSTTATPTPPPTSDYKTAQTAYDT 120
DB 61 lrsavnlmsladklgltassnsstsrsvdvtatptpppttdyktqactaydt 120
QY 121 IFTSTSLADIQAALVSLQDAVTNPKDAADETAIAAEWETKNADAIKVGAQITELAKY 180
DB 121 iftstsladiqaalvslqdvtnpkdtaadeetaiaaewetknadavkvgaqitelaky 180
QY 181 ASDNQAILDLSGLKLTSPDLLQALQSANNKKAELLKEMQDNPVPGKTPAQAQSLVD 240
DB 181 asdnqaildlsigkltstfllqallqsvannnkkaellkemqdnvpgktpaqaqslvd 240
QY 241 QTDATATQIEKDGNAIGDAYFAGNAGAVENAKSNNSISNIDSAKAAIAATAKQIAEAQ 300
DB 241 qtdatatqiekdgnaigdayfagnagavenaksnnsisnidsakaaiaataktqieaq 300
QY 301 KKFDPSPILQAEQMVIOAEKDLKNIPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMML 360
DB 301 kkfpdspilqaeqmvioaeqkdlknipadgsdvpnpgtttvggskqggssigsirvsmll 360
QY 361 DDAENETASILMSGFROMIHMFENTENPDSQAQQLAQAARAAKAGDDSAALADAQK 420
DB 361 ddaenetasilmsgfromihmfentpdsqaqqlaqaaraakagddsaaladaqk 420
QY 421 ALEALGRAGQQGTLNALGQIASAAVYAGVLPQQVLTWRARYQAYVEQKLISEEDLN 480
DB 421 alealgragqqgtlnalqgiassaaavyagvlpqqvltwiraryqayveqkliseedln 480
QY 481 SAVDHHHHH 490
DB 481 savdhhhhh 490

RESULT 2
AAY71954
ID AAY71954 standard; Protein; 651 AA.
XX
AC AAY71954;
XX
DT 26-MAR-2001 (first entry)

```

```

XX Chlamydia pneumoniae 76 kDa full-length protein.
XX DE
XX KW 76 kDa protein; bactericidal; diagnosis; prevention; treatment;
XX KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
XX KW acute respiratory disease; cough; sore throat; hoarseness; fever;
XX KW vaccine; immunisation.
XX OS Chlamydia pneumoniae.
XX PN WO200066739-A2.
XX PD 09-NOV-2000.
XX PF 03-MAY-2000; 2000WO-CA00511.
XX PR 03-MAY-1999; 99US-0132270.
XX PR 30-JUN-1999; 99US-0141276.
XX PA (AVET ) AVENTIS PASTEUR LTD.
XX PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX DR WPI; 2000-687542/67.
XX DR N-PSDB; AAD02063.
XX PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX PT useful for vaccinating against Chlamydia infections -
XX PS Claim 16a; Fig 1; 112pp; English.
XX CC The present sequence is Chlamydia pneumoniae full-length
XX CC 76 kDa protein. C. pneumoniae 76 kDa protein
XX CC is used in the diagnosis, prevention and treatment
XX CC of C. pneumoniae infections (e.g. pneumonia, upper respiratory
XX CC tract disease, bronchitis, sinusitis and acute respiratory
XX CC disease such as cough, sore throat, hoarseness, fever; and
XX CC abnormal chest sounds on auscultation). C. pneumoniae sequence
XX CC is also used as vaccines for immunising humans against diseases
XX CC caused by C. pneumoniae.
XX CC
XX SQ Sequence 651 AA;

Query Match 90.7%; Score 2188; DB 21; Length 651;
Best Local Similarity 98.9%; Pred. No. 4.3e-139;
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MWNPIGPGPIDETERTPPADLSAAGLEASAAKSAEQAQRIAGAAKPKESKTSVERWSI 60
DB 1 mwnpiggpiddetertppadlsaggleasaanksaeeqriagaekpkesktdsverwsi 60
QY 61 LRSVAVNLMSTADKLGITASSNSSTSRSDVDSTTATPTPPPTSDYKTAQTAYDT 120
DB 61 lrsavnlmsladklgltassnsstsrsvdvtatptpppttdyktqactaydt 120
QY 121 IFTSTSLADIQAALVSLQDAVTNPKDAADETAIAAEWETKNADAIKVGAQITELAKY 180
DB 121 iftstsladiqaalvslqdvtnpkdtaadeetaiaaewetknadavkvgaqitelaky 180
QY 181 ASDNQAILDLSGLKLTSPDLLQALQSANNKKAELLKEMQDNPVPGKTPAQAQSLVD 240
DB 181 asdnqaildlsigkltstfllqallqsvannnkkaellkemqdnvpgktpaqaqslvd 240
QY 241 QTDATATQIEKDGNAIGDAYFAGNAGAVENAKSNNSISNIDSAKAAIAATAKQIAEAQ 300
DB 241 qtdatatqiekdgnaigdayfagnagavenaksnnsisnidsakaaiaataktqieaq 300
QY 301 KKFDPSPILQAEQMVIOAEKDLKNIPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMML 360
DB 301 kkfpdspilqaeqmvioaeqkdlknipadgsdvpnpgtttvggskqggssigsirvsmll 360
QY 361 DDAENETASILMSGFROMIHMFENTENPDSQAQQLAQAARAAKAGDDSAALADAQK 420

```



```

Db 361 ddaenetasilmgrfmhmtentpdsqaaqgelaagaaaraakaagddsaaladagk 420
Oy 421 ALEAALGRAGQOQOGLNALGOIASAAVVSAGVLP 454
Db 421 aleaalgragqoqoqglnalqiasaavvsagvpp 454

RESULT 3
AAV71956
ID AAV71956 standard; Protein; 452 AA.
XX
AC AAV71956;
XX
DT 26-MAR-2001 (first entry)
XX
DE 3'-truncated Chlamydia pneumoniae 76 kDa protein.
XX
KW 76 kDa protein; bactericidal; diagnosis; prevention;
KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
KW acute respiratory disease; cough; sore throat; hoarseness; fever;
KW vaccine; immunisation; treatment; truncation mutant; mutain.
XX
OS Chlamydia pneumoniae.
OS Synthetic.
XX
PN WO200066739-A2.
XX
PD 09-NOV-2000.
XX
PF 03-MAY-2000; 2000WO-CA00511.
XX
PR 03-MAY-1999; 99US-0132270.
PR 30-JUN-1999; 99US-0141276.
XX
PA (AVET ) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI; 2000-687542/67.
XX
DR N-PSDB; AAD02065.
XX
PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
PT useful for vaccinating against Chlamydia infections -
XX
PS Claim 16c; Page 104-106; 112pp; English.
XX
CC The present sequence is 3'-truncated
CC Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
CC protein is used in the
CC diagnosis, prevention and treatment of C. pneumoniae
CC infections (e.g. pneumonia, upper respiratory
CC tract disease, bronchitis, sinusitis and acute respiratory
CC disease such as cough, sore throat, hoarseness, fever; and
CC abnormal chest sounds on auscultation). C. pneumoniae sequence
CC is also used as vaccines for immunising humans against diseases
CC caused by C. pneumoniae.
XX
SQ Sequence 452 AA;

Query Match 90.5%; Score 2184; DB 21; Length 452;
Best Local Similarity 99.1%; Pred. No. 5e-139;
Matches 448; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 MWNPIGPGPIDETPTPADLSAOGLEASAANKSAEAOIRAGAEAKPKESKTDTSVERWSI 60
Db 1 mwnpi9ppidetertppadlsaqgleasaanksaegriagaeakpkcsdsvrws1 60
Oy 61 LRSAYNALMSLADKLGIASNSSSSSTRSADVDSTTATPTPPPTSDYKTQATAYDT 120
Db 61 lrsavnalmsladkligiassnssstrsadvdsttatptppptfddyktqataydt 120

```

```

Oy 121 IFTSTSLADIAALVSLQDAVTNIKDTAATDEETAIATAEWETKNADAIVKGAQITELAKY 180
Db 121 iftstsladiqaalvslqdavtnikdaatdeetaiaaewetknadavkvgaqitelaky 180
Oy 181 ASDNQAILDSLGKLTSLFDLLQTLALLOSVAANNKAAELLKEMQDNPPVVPKTTAIAQSLVD 240
Db 181 asdnqaildsilgkltslfdllqallqsvannnkaeallkemqdnppvvpktpaiadslvd 240
Oy 241 QTDATATOIEKDCNAGIDAYFAGONASGAVENAKSNNSISNIDSAKAAIATATAKTQIAEAQ 300
Db 241 qtdatatqiekdgdnairdayfagqnasgavenaksnsinsidsakaatataktqiaeq 300
Oy 301 KKFPDSPILQEAEMVIOAEKDLKNIKPADGSDVPNPPTTGGSKQOQSSIGSIRVSMML 360
Db 301 kkfpdspilqeaemviqueakdlknikpadgsdvpnppttvggskqgssigsirvsmll 360
Oy 361 DDAENETASILMSGFROMIHMENTENPDQAAQOELAAQARAARAKAGDDSAALADAQK 420
Db 361 ddaenetasilmgrfmhmtentpdsqaaqgelaagaaaraakaagddsaaladagk 420
Oy 421 ALEAALGRAGQOQOGLNALGOIASAAVVSAGV 452
Db 421 aleaalgragqoqoqglnalqiasaavvsagv 452

RESULT 4
AAV35358
ID AAV35358 standard; Protein; 478 AA.
XX
AC AAV35358;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed polypeptide.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST ) GENSET.
XX
PI Griffiths R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1156; Disclosure; 1912pp; English.
XX
CC AAV34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 478 AA;

```



Query Match	89.6%	Score	2162;	DB	20;	Length	478;
Best Local Similarity	99.1%;	Pred. No.	1.6e-137;				
Matches	443;	Conservative	1;	Mismatches	3;	Indels	0;
Gaps	0;						
QY	1	MVNPIGPGPIDETERTPPADLSAQGLSASANKGAEAQRIAGACAKPKESKTDVSVERWSI	60				
DB	3	mvnpigppideterppadlsaggleasaankaesaqriagaaekpkesktidverwsi	62				
QY	61	LSRAVNALMSLADKLGIASNSSSSTGRSDVDSTTATPTPPPTSDYKTKQATAYDT	120				
DB	63	lrsavnalmsladklgiassnsstsrdsadvdsttatptppptfdyktqagtaydt	122				
QY	121	IFTSTSLADTQAALVLSQDQAVTNIKDTAAATDEETFAIAAETKNADAKVGAQITELAKY	180				
DB	123	iftstsladiqaalvlsiqdavtnikdtaadeetaaewaetknadavkvgaqitelaky	182				
QY	181	ASDNQAILSDSGKUTSFDLLQTLALLQSVANNKAAELLKEMQDNVPVPGKTPATAQSLVD	240				
DB	183	asdnqailsdsgkutsfdllqalqlsvannnkaaellkemqdnvpvpgktpaiagslvd	242				
QY	241	QTDATAQIEKQGNAGIDGAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTQIAEAQ	300				
DB	243	qtdatatqiekdgnaidgayfagqnasgavenaksnnsidnsidakaaiataktqiaeaq	302				
QY	301	KKPPDSPILQAEQMWTOAEKDKLNKIPAGDSVPNPGTTVGGSKQOGSSISIRVSMLL	360				
DB	303	kkfpdspilqaeaqmviqaeekdklnkipagdsvpnpgttvggskqgssisirvsmll	362				
QY	361	DDAENETASTILMSGFROMIHMFNTENPDSAAQOELAQAARAKAAGDSDASAAALADAQK	420				
DB	363	ddaaenetasilmgfromihmfntcnpdsaaqqaelaqaaraakaagdsdaaaaladaqk	422				
QY	421	ALEAALGKAGQOQOILNALGQIASAAY	447				
DB	423	aleaalgkagqggqilnalqgiasaav	449				

RESULT	5	
AAAY1955		
ID	AAAY1955 standard; Protein; 583 AA.	
XX		
AC	AAAY1955;	
XX		
DT	26-MAR-2001 (first entry)	
XX		
DE	5'-truncated Chlamydia pneumoniae 76 kDa protein.	
XX		
KW	76 kDa protein; bactericidal; diagnosis; prevention;	
KW	pneumonia; upper respiratory tract disease; bronchitis; sinusitis;	
KW	acute respiratory disease; cough; sore throat; hoarseness; fever;	
KW	vaccine; immunisation; treatment; truncation mutant; mutein.	
XX		
OS	Chlamydia pneumoniae.	
OS	Synthetic.	
XX		
FT	Key	Location/Qualifiers
FT	Misc-difference 497	/note= "Encoded by ACT"
XX		
XX	WO200066739-A2.	
XX		
PD	09-NOV-2000.	
XX		
PF	03-MAY-2000; 2000WO-CA00511.	
XX		
PR	03-MAY-1999; 99US-0132270.	
PR	30-JUN-1999; 99US-0141276.	
XX		
PA	(AVET ) AVENTIS PASTEUR LTD.	
XX		
PI	Murdin AD, Oomen RP, Wang J, Dunn P;	

```

XX WP1; 2000-687542/67.
XX N-PSDB; AAD02064.
XX
XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX useful for vaccinating against Chlamydia infections -
XX
XX Claim 16b; Page 100-102; 112pp; English.
XX
XX The present sequence is 5'-truncated
XX Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
XX protein is used in the
XX diagnosis, prevention and treatment of C. pneumoniae
XX infections (e.g. pneumonia, upper respiratory
XX tract disease, bronchitis, sinusitis and acute respiratory
XX disease such as cough, sore throat, hoarseness, fever; and
XX abnormal chest sounds on auscultation). C. pneumoniae sequence
XX is also used as vaccines for immunising humans against diseases
XX caused by C. pneumoniae.
XX
XX Sequence 583 AA;
XX
Query Match 76.7%; Score 1850; DB 21; Length 583;
Best Local Similarity 98.7%; Pred. No. 2e-116;
Matches 381; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 69 MSADKLGIASSSSSSRSADVSTATAPTPPPPTSDYKTQATAYDTIFTTSLA 128
Db 1 msldkligiassssssrsadvstataptppptfddykttgactydtiftstla 60
QY 129 DIOAALVSLQDVTNLIKDTAATDEEPAIAAEWETKNADATKVGAOITELAKYASDNOAIL 188
Db 61 dfgaalvslqdvtnlikdtaadeetaiaaewetknadavkvgaqitelaakysdngail 120
QY 189 DSIQKLTISFDLLQALQSVANNKKAELKEMQDNVPVPGKTPAIAQSLVDQTDATATQ 248
Db 121 dsiqkltisfllqaalqsvannkkaaelikemqdnvpvpgktpaiaqslvdqtda:atq 180
QY 249 IEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSKAAIAIAKTQIAEAQKKFPDSPI 308
Db 181 iekdgnaigrdayfagnasgavenaksnnsisnidskaaiaiaaktqiaeaqkkfpdsp 240
QY 309 LQAEQMVIOAEKDLKNIPKPADGSDVPNPQTTVGGSKQCGSSIGSIRVSMLLDDAENETA 368
Db 241 lqaeqmvioaekdlknipkpadgsdvnpnpqttvggskqcgssigsvrsmllddae:eta 300
QY 369 SILMSGFQRMHFWNTENPDSQAQBELAAQARAANKAGDDSAALADAKALEAALGK 428
Db 301 silmsgrqmihmfnntenpdsqaqbelaaqaraakaagddsaaladagkaaleaalgk 360
QY 429 AGQQQGLINALGQTASAAVVSAGVLP 454
Db 361 agqqqgilnalqiasaavvsagvpp 386
RESULT 6
AAY37571
ID AAY37571 standard; Protein; 350 AA.
XX
XX AAY37571;
AC
XX
XX
XX 07-OCT-1999 (first entry)
XX
XX Chlamydia trachomatis surface exposed protein.
XX
XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
XX nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX Bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
XX Chlamydia trachomatis.
XX

```



PN W09928475-A2.  
 PD 10-JUN-1999.  
 XX 27-NOV-1998; 98WO-IB01939.  
 XX 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97FR-0015041.  
 PR 17-DEC-1997; 97FR-0016034.  
 XX (GEST ) GENSET.  
 XX Griffais R;  
 PI WPI; 1999-371125/31.  
 DR  
 XX  
 XX Genome sequence of Chlamydia trachomatis  
 PT  
 XX Disclosure; Page 1226-1227; 1755pp; English.  
 XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see A201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perihpatitis, Bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 XX  
 XX Sequence 350 AA;

Query Match 11.7%; Score 281.5; DB 20; Length 350;  
 Best Local Similarity 24.9%; Pred. No. 2.6e-11;  
 Matches 95; Conservative 71; Mismatches 122; Indels 93; Gaps 13;  
 QY 2 VNPIGPGPIDETERTPPADLSAOGLEASAAKSAEAOIAEAKPKESKTSVERWSIL 61  
 Db 23 mnpingqi-----asnetkestasea-----ssassssvswfl 60  
 QY 62 RSAYNALMSLADKLGIASSNSSTSRSDVDSTTATPTPPPTSDDYKTAQTAYDTI 121  
 Db 61 ssaknalislrd--ailnknssptdsis-qleatsts-tvtrvaakdy-dkaksnfda 115  
 QY 122 FT-----STSLADIOAALVSLQ-----DAVTNKDTAATDEETAIAAEW 160  
 Db 116 ksglenaktlaeyletkmadlmaalqdmansdpndhteelnikkaleaqkdtl----- 170  
 QY 161 ETKNADAIKVGAOITELAKYASDNOAILDSLGKLTSPFLOLTALLOSVAANNKAEELLE 220  
 Db 171 -----dkinklvtnqknslteakltkdsadqipainrlleinknsahqilke 219  
 QY 221 MODNPVPGKTPAIQAISLVDTATATQIEKD-----GNAIGDAYFAGONASGA 269  
 Db 220 lkeq-----isnykavltdevkivefseagiklgalqsvdaggqsgaa 265  
 QY 270 VENAKSNNSISNIDSAKAAITAKTQIAFAO---KKFPDPSILOAEQOMVTOAEKDLNI 326  
 Db 266 vlqarqnsnpdniaatklidaaktvnelkqehqeadspdvkkaeqinqagqdiqti 325  
 QY 327 KPAGSDVPNPGTTVGGSKQQ 347  
 Db 326 tps-gldipivgpgsgxpqe 345

RESULT 7  
 AAB13695  
 ID AAB13695 standard; Protein; 361 AA.  
 XX  
 AC AAB13695;

XX 02-FEB-2001 (first entry)  
 DT Chlamydia sp. protein # 6.  
 XX  
 DE Chlamydia sp.  
 KW Chlamydial infection; sexually transmitted disease;  
 KW pelvic inflammatory disease; PID; tubal obstruction; infertility;  
 KW trachoma; blindness; acute respiratory tract infection;  
 KW atherosclerosis; coronary heart disease; antibacterial.  
 XX  
 OS Chlamydia sp..  
 XX WO2000034483-A2.  
 PN 15-JUN-2000.  
 PD  
 XX 08-DEC-1999; 99WO-US29012.  
 PF 08-DEC-1998; 98US-0208277.  
 PR 08-APR-1999; 99US-0288594.  
 PR 01-OCT-1999; 98US-0410568.  
 PR 22-OCT-1999; 99US-0426571.  
 XX (CORI-) CORIXA CORP.  
 XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;  
 XX WPI; 2000-431303/37.  
 DR  
 XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection  
 XX comprises immunogenic portion of Chlamydia antigen, which comprises  
 PT amino acid sequence encoded by polynucleotide sequence -  
 PT  
 XX Disclosure; Pages 251-253; 256pp; English.

The present invention relates to new nucleic acid sequences and the  
 CC proteins encoded by the nucleic acid sequences. The encoded proteins  
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded  
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia  
 CC infection. Chlamydiae are intracellular bacterial pathogens that are  
 CC responsible for a wide variety of human infections. C. trachomatis  
 CC infection is one of the most common sexually transmitted diseases and can  
 CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
 CC and infertility. Trachoma due to ocular infection with C. trachomatis is  
 CC the leading cause of preventable blindness worldwide. C. pneumoniae is a  
 CC major cause of acute respiratory tract infections in humans and is also  
 CC thought to play a role in the pathogenesis of atherosclerosis and  
 CC coronary heart disease. The present sequence is a protein isolated in the  
 CC present invention.  
 XX Sequence 361 AA;

Query Match 9.6%; Score 231.5; DB 21; Length 361;  
 Best Local Similarity 35.4%; Pred. No. 6.3e-08;  
 Matches 57; Conservative 30; Mismatches 61; Indels 13; Gaps 5;  
 QY 301 KKFPDPSILOAEQOMVIOAEKDLNKPADGSDVPNPGTTVGGSKQSSSTGSI----- 354  
 Db 2 qeiadspvlkkaeqinqagqdiqtips-gldipivgpgs-gsaasagaaalkssnns 59  
 QY 355 -RVSMMLDDAENETASILMSGFROMIHMFNTENPDSSAAQOELAAQARAAK---AAGDD 410  
 Db 60 grisllldvdmamaaiaamqgfrsmieqfvrnnpatakelameqltamsdqlvgadg 119  
 QY 411 AAALADAOAKALEALGKAGQOQGLNALGQTASAAVVSAG 451  
 Db 120 lpaelqaiakdalqal-kpsadglatamgvaafaaakvg 159

RESULT 8  
 AAG83263  
 ID AAG83263 standard; Protein; 361 AA.







Db 1072 iddnpnltpde-----kesaknaveeaaakvata-aidkastpdavqveedkgv--aaain 1122  
 Qy 67 ALMSIADKLGITASSNSSSTSRAD-----VDSTATAPTPTPPPSDDYKTKQAOTAY 118  
 Db 1123 litakadagvialadeiklkedkqaeakaiaasmt-----neekalakkal 1173  
 Qy 119 DTIFTSTSLADIAALVSLODA---VTN- IKDTAATDETAIAAEWETKNADAIVKGAQI 174  
 Db 1174 qdv-----vdkgkaeledaarvatneiheahttekakaealageksltdtkearda 1225  
 Qy 175 TELAKYASDNQAILDSLCK--LTSEFLLQPTALLQSVANNKAAELLKEMQNPVVPK-- 230  
 Db 1226 velak---dke-----igkeairteeeeatkiveklaedtr-----kaiednplnsdedk 1272  
 Qy 231 -----TPALAQSLV---DOTDATATQIEKDGNAGIDAYFAGQNASGAVENAKSNNSIS 280  
 Db 1273 qaeikkltdavaktlatmndnadrtdgaek-aqalad-----lekaketqkia 1320  
 Qy 281 NIDSAKAI-----AT---AKTOIAE-----AOKKFPDPSPILOPAEQMVIOAE 320  
 Db 1321 d---kaaidr-ltilvkdgeleackqaktkiaakdaaaakeaiaasnplntdaekktftda 1376  
 Qy 321 KDLNIRPADGSDVPNCTTTVGSKQOGSSIGSIRVSM---LDDAENETASILMSGRFQ 377  
 Db 1377 vdaevakandaiaatspadv--qkeedagvaaiaedvldaaqkdaknkiaakdaaaa-ke 1433  
 Qy 378 MIHMFNTENPDPSQAAQQL-----AAQARAAKAAGDDSAALADAKKALEAALGRAGQOOG 434  
 Db 1434 ai-----gsnplntdaekktftdavaevakandaiaatspadvqkeeda--gvaiaaed 1487  
 Qy 435 ILNALGQIASAAVSVAGVLPLOQVILIRYQVYVVEOKLISEED-LNSAVD 484  
 Db 1488 vldaaqkda-----knkiaesdaaksaid 1512

RESULT 10  
 AAY37572  
 ID AAY37572 standard; Protein; 331 AA.  
 AC AAY37572;  
 XX  
 XX  
 XX 07-OCT-1999 (first entry)  
 DE Chlamydia trachomatis cellular envelope protein.  
 XX  
 XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihemipatits;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
 XX Chlamydia trachomatis.  
 OS  
 XX W09928475-A2.  
 PN  
 XX 10-JUN-1999.  
 PD  
 XX 27-NOV-1998; 98WO-IB01939.  
 PF  
 XX 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97FR-0015041.  
 PR 17-DEC-1997; 97FR-0016034.  
 XX  
 XX (GEST ) GENSET.  
 PA  
 XX Griffiths R;  
 PI  
 XX WPI; 1999-371125/31.  
 DR  
 XX Genome sequence of Chlamydia trachomatis  
 PT  
 XX Disclosure; Page 1227-1228; 1755pp; English.  
 PS  
 XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC

CC of Chlamydia trachomatis (see AA0201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perihemipatits, bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.

XX Sequence 331 AA;

Query Match 6.8%; Score 164.5; DB 20; Length 331;  
 Best Local Similarity 38.5%; Pred. No. 0.0018;  
 Matches 45; Conservative 12; Mismatches 53; Indels 7; Gaps 3;

Qy 338 GTTVGSKQOGSSIGSIRVSMMLDDAENETASILMSGRFMIHMFNTENPDPSQAAQQLA 397  
 Db 17 gsaagalk---ssnsggrisllldvndemaaiamqgfirmsieqfnvnpnatakelqame 73  
 Qy 398 AQARA---AGDDSAALADAKKALEAALGRAGQOOGILNALGQIASAAVSVAG 451  
 Db 74 aqltamsdqlvgadgelpaeigaiakdalqal-kqpsadglatamngqvafaaakvgg 129

RESULT 11

AAY81609  
 ID AAY81609 standard; Protein; 1237 AA.

AC AAY81609;

XX  
 XX 24-MAY-2000 (first entry)

XX Streptococcus pneumoniae type 4 protein sequence #109.

DE Streptococcus pneumoniae; vaccine; screening; protein antigen;  
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
 KW pneumococcal disease.

XX Streptococcus pneumoniae.

XX W0200006737-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

XX 19-MAR-1999; 99US-0125164.

XX (MICR-) MICROBIAL TECHNIQS LTD.

XX Gilbert CFG, Hansbro PM;

XX WPI; 2000-195300/17.

XX New Streptococcal protein, useful as a vaccine, for diagnosis of  
 PT pneumococcal diseases and for screening agents capable of antagonizing  
 PT or inhibiting expression of the protein

XX Claim 1; Page 83; 108pp; English.

XX AAY81501 to AAY81679 represent specifically claimed protein sequences  
 CC isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent  
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.  
 CC The sequences have antibacterial and antiinflammatory properties.  
 CC The protein sequences, and fragments of them, are useful as immunogens  
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in  
 CC diagnostic assays. The proteins and nucleotides can be useful for the  
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also  
 CC useful for screening an agent capable of antagonising, inhibiting or



CC interfering with the function or expression of the proteins in which the  
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection  
 CC and meningitis. AA05591 to AA05614 represent primers used in the  
 CC exemplification of the present invention.

XX Sequence 1237 AA;

Query Match 6.4%; Score 153.5; DB 21; Length 1237;  
 Best Local Similarity 19.9%; Pred. No. 0.054;  
 Matches 84; Conservative 84; Mismatches 205; Indels 49; Gaps 9;

QY 13 TERPPADLSAQGLEASANKSAEQAQRIAGAEAKPKESKTDSEVERWSILRSVAVNMLMSLA 72  
 Db 467 tsasasastsasasastsasasastsasasastsasasastsasasastsasasastsas 526  
 QY 73 DKLGIASSNSSSSRSADVDSTTATPTPPPTSDYKTAQATYDTITFTSLADIQA 132  
 Db 527 ---gsaststasastsasasastsasasastsasastsasastsasastsasastsasast 577  
 QY 133 ALVSLQDAVNIKDTATDEETATAAEWETKNADAIVKGAQITELAKYASDNOAILDLSLG 192  
 Db 578 stsasastsasast---sasasastsasasastsasasastsasasastsasastsasasts 634  
 QY 193 KLTFDLLQTLQSVANNKKAELLKEMODNPVPGKTPAIAQSLVDQTDATATQIEKD 252  
 Db 635 asasastsasasastsasast-----sasastsasasastsasasastsasastsas 686  
 QY 253 GNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAK-ATAATKTAQIAEAQKFPDPILOE 311  
 Db 687 astsasastsasasastsasastsasastsasastsasastsasastsasastsasastsa 741  
 QY 312 AEQWVIAQEKDLKNIKPADGSDVPNPGTGVGSKQGGSSI-GSTRVSMILLDDAENETASI 370  
 Db 742 -----sastsasastsasastsasastsasastsasastsasastsasastsasast 785  
 QY 371 LMSGFRQIMHFMNTENPDPSQAQOELAAQARAAGDSDSAAALADAQAALGKAG 430  
 Db 786 sas-----esastsasaststasastsasastsasastsasastsasastsasastsa 839  
 QY 431 QQ 432  
 Db 840 iq 841

RESULT 12  
 AAG91443  
 ID AAG91443 standard; Protein; 600 AA.  
 AC AAG91443;  
 XX  
 XX  
 XX 26-SEP-2001 (first entry)  
 XX C glutamicum protein fragment SEQ ID NO: 5197.  
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 XX Corynebacterium glutamicum.  
 XX  
 PN EPI108790-A2.  
 XX  
 XX 20-JUN-2001.  
 PD  
 XX  
 XX 18-DEC-2000; 2000EP-0127688.  
 PF  
 XX  
 XX 16-DEC-1999; 93JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 XX WPI: 2001-376931/40.  
 DR N-PSDB; AAH66662.  
 XX  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX Claim 17; SEQ ID NO: 5197; 246pp + Sequence Listing; English.  
 XX  
 XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: the sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 XX Sequence 600 AA;

Query Match 6.3%; Score 153; DB 22; Length 600;  
 Best Local Similarity 21.6%; Pred. No. 0.023;  
 Matches 96; Conservative 66; Mismatches 187; Indels 96; Gaps 14;

QY 18 PADLSAQGLEASANKSAEQAQRIAGAEAKPKESKTSVE-RWSILRSVAVN-ALMSLADKL 75  
 Db 24 pddaaiaqaeevsgdgvargiags-lsstdaeinrvelmgaltreevksivldhdaq 82  
 QY 76 GIASSNSSSTSRSDVDSTTATPTPPPTSDYKTAQATYDTITFTSLADIQALV 135  
 Db 83 alae-----qagdalaakkdldsgaqie 107  
 QY 136 SLQDAVNIKDTATDEETATAAEWETKNADAIVKGAQITELAKYASDNOAILDLSGLT 195  
 Db 108 aagerldeisraayrqngtskglsgisngnsedalrqtlyrtsaekqqaaveelrlr 167  
 QY 196 SFDLLQTLQSVANNKKAELLKEMODNPVPGKTPAIAQSLVDQTDATATQIEKDGA 255  
 Db 168 t-----enaukesvliq-----arivaegreaveekqvqea 200  
 QY 256 IGDYFAGQNASGAVENAKSNNSISNIDSAAKATATATAKTQIAEAQKFPDPILOEAEOM 315  
 Db 201 ---aiaanseqlnvltnrs-tlvagrdgaernlalaradadnlqgdraeeyefqgaqa 256  
 QY 316 VIOAEKD----LKNIKPADGSDVPNPGTGVGSKQGGSSIGSIRVSMILLDDAENETASIL 371  
 Db 257 riqaeeaaqaeeekrradea-aaqaaeeaaqaeeaaqaeeaaqaeeaaqaeeaaqa- 314  
 QY 372 MSGFRQIMHFMNTENPDPSQAQOEE-----LAAQARAAGDSDSAAALADAQAALGK 425  
 Db 315 -----aetqa 362  
 QY 426 LGKAGQQOQILNALGQIASAAVNSA 450  
 Db 363 -aqalreqaltaa--saaaaallaa 384

RESULT 13  
 AAW02098  
 ID AAW02098 standard; Protein; 1561 AA.  
 XX  
 XX AAW02098;  
 AC  
 XX  
 XX 27-OCT-1996 (first entry)



XX DE S. mutans antigen I/II.  
 XX KW Caries; antigen I/II; epitope; tooth decay; vaccine; immunisation.  
 XX OS Streptococcus mutans.  
 XX FH Key Location/Qualifiers  
 XX FT Misc-difference 618..650  
 XX FT /note= "amino acids 618-650 differ from the  
 XX FT residues deduced from the nucleic acid  
 XX FT sequence."  
 XX PN WO9623886-A1.  
 XX PD 08-AUG-1996.  
 XX PF 31-JAN-1996; 96WO-GB00207.  
 XX PR 31-JAN-1995; 95GB-0001826.  
 XX PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 XX PI Kelly C, Lehner T;  
 XX DR WPI; 1996-371434/37.  
 XX DR N-PSDB; AAT36122.  
 XX KW Nucleic acid encoding polypeptide for prevention or treatment of  
 XX PT dental caries - which stimulates T or B cell response, and/or  
 XX PT adheres to tooth in competition with Streptococcus mutans antigen  
 XX PT I/II  
 XX PS Disclosure; Page 45-46; 63pp; English.  
 XX CC Streptococcus mutans antigen I/II (AAW02098) is a 185 kDa cell surface  
 XX CC protein at least partly responsible for S. mutans adhesion to teeth.  
 XX CC The I/II antigen includes a series of overlapping T-cell, B-cell and  
 XX CC adhesion epitopes. Fragments (see also AAT36111-21) of the I/II  
 XX CC antigen gene (see also AAT36122) can be used to produce recombinant  
 XX CC polypeptides (AAW02087-97) carrying such epitopes for use in vaccines  
 XX CC for immunisation against dental caries.  
 XX SO Sequence 1561 AA;  
 Query Match 6.3%; Score 152; DB 17; Length 1561;  
 Best Local Similarity 22.3%; Pred. No. 0.093;  
 Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;  
 Qy 48 KESKTSVERWSILRSVAVNALMSLADKLGIASSN-SSSSTSRKADVDSTTATPTPPPT 106  
 Db 2 kvktygrkskiskitlgavlgvtvaavsvagkvkfadetttsdvdkvvgvtgtnpat 61  
 Qy 107 -----SDDYKTKQAQAYDTIFTST-----SLADIQAALVS-LQDVTNINIKDT 147  
 Db 62 nlpeaqsasqaeqsktqlerqmvhtievkptldlqaaakdaksagvnnvqadvn-kgt 120  
 Qy 148 AATDEETAIAREWE-----TKNADAIK-----VCAQITELAKYASNDQAILDSLKG 193  
 Db 121 vktae-avqketeikedytkqaeidkkttdqykdsvaaheaevakikaknqatkeqygk 179  
 Qy 194 LTSFDLL-QTALLQSVANNKAAEL-----LKEMQDNVPVPGKYPATIAQSLVDQTDATAT 247  
 Db 180 ----dmvahkaeverinaaasktayeklaqyqadlaavqktnaanqasy-qkalaay 234  
 Qy 248 QIE-----KGNAGIDAYFAGNAGAVENAKSNNSISNIDSAKAIATAK-----TQIAEA 299  
 Db 235 qaelkrvqeanaaakaaqy---dtavaannakntelaanaeairkrnakataeyetklaqy 291  
 Qy 300 Q---KKFPDPSILOFAE-QWVIOA-EKDLKNKPAD-----GSDVPNPGTTVG 342  
 Db 292 qaelkrvqeananeadyqakityqtelarvqkanadakaayaevaannaknaaltcae 351

Qy 343 GS-----KOGSSIGSIRVSMLLDDAENETASTILMSGFRQTHMFNTE 385  
 Db 352 ntaikqrnenakatyaaalkyeadlaavkkanaaneadyqak---ltaqtelarvqka 408  
 Qy 386 NPDQAQOQELAAQARAQAA--GDSSAAAAL-ADAQKALEAALGK 428  
 Db 409 nedakaayeaavaanaanaaltaentaikkrnadakadyeaklak 454  
 RESULT 14  
 AAR84568  
 ID AAR84568 standard; Protein; 643 AA.  
 XX AC AAR84568;  
 XX DT 09-MAY-1996 (first entry)  
 XX DE Trypanosoma cruzi TCR27 polypeptide, Ag15.  
 XX KW TCR27; Chagas disease; repeat unit; diagnosis; blood screening;  
 XX KW recombinant; fusion protein; glutathione-S-transferase.  
 XX OS Trypanosoma cruzi.  
 XX FH Key Location/Qualifiers  
 XX FT 329..552  
 XX FT /label= repeat\_region  
 XX FT /note= "16 of 69 repeat units of 14 amino acids"  
 XX PN WO9525797-A1.  
 XX PD 28-SEP-1995.  
 XX PF 20-MAR-1995; 95WO-US03191.  
 XX PR 24-MAR-1994; 94US-0216894.  
 XX PA (KIRC/) KIRCHHOFF L V.  
 XX PA (OTSU/) OTSU K.  
 XX PI Kirchoff LV, Otsu K;  
 XX DR WPI; 1995-344618/44.  
 XX DR N-PSDB; AAT05332.  
 XX PT New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi - as  
 XX PT immunoassay reagent for specific diagnosis of Chagas disease, also  
 XX PT related nucleic acid and transformed cells  
 XX PS Disclosure; Page 40-41; 68pp; English.  
 XX CC AAR84565-R84569 are polypeptides of the TCR27 protein of T.cruzi  
 XX CC The proteins are all fusion products with glutathione-S-transferase  
 XX CC (GST) and some contain a linker sequence. The TCR27 protein comprises  
 XX CC a 95 amino acid (aa) N-terminal region; 69 repeats of a highly  
 XX CC conserved 14 aa sequence and a 68 aa C-terminal region. This sequence  
 XX CC encodes the GST sequence, the Ag44 polypeptide contg. 16 of the 69  
 XX CC repeat units and also contains the amino and carboxy terminal  
 XX CC peptides of TCR27. The TCR27 polypeptides of the invention are useful  
 XX CC for the diagnosis of Chagas disease (American Trypanosomiasis), they  
 XX CC are capable of detecting anti-T.cruzi antibodies; or for blood  
 XX CC screening. The TCR27 protein has epitopes to which most T.cruzi  
 XX CC infected individuals have antibodies. The TCR27 polypeptides will not  
 XX CC react with serum from patients with leishmaniasis, schistosomiasis,  
 XX CC or autoimmune disease and are hence less likely to cause false  
 XX CC positives in diagnosis.  
 XX SO Sequence 643 AA;

Query Match 6.2%; Score 150.5; DB 16; Length 643;  
 Best Local Similarity 21.5%; Pred. No. 0.037;



Matches	99;	Conservative	58;	Mismatches	169;	Indels	135;	Gaps	17;
QY	7	PGPIDETERTPADLSAAGLEASAAKSAEAQRIAGAEAKPKESKTDTSVERVSILRSAVN	66						
Db	216	pkpsdlvrgpsql--qqaennitnskkmtklrekvkaekikda inratkleeern	273						
QY	67	ALMSLDLKLGIASSNSSSTSRSDVDSTTATPTPPPTSDDYKTAQATAYDTIFTSTS	126						
Db	274	qaykaahk---aeekaktfgrlftfeseninlkkrp-----	307						
QY	127	LADIQAALVSLQDAVTN-----IKDTAATDE-----ETAIAAEWETKNADAIKVGAQI	174						
Db	308	-----ndavsnrdrkknsctaktdevekgraaeaaakavetekgra-----aea	350						
QY	175	TELAKYSDNOAILDSLGLKTSFOLLQTALLQSYANNKAAELLKEMQDNVPVPGKTPAI	234						
Db	351	tkvae-----aekrikaeaaakavetekgraaeatkv	381						
QY	235	A-----QSLVDQTDATATOIEKDGNAIGDAYFAGONASGAVENAKSN-----NSTISNTD	283						
Db	382	aaekqkaaeaaakavetekgraaeatkvaaekqgraaeamkvaeeekqkaaeatkvaee	441						
QY	284	SAKAIAIATAKTOIAEPAQR-KFPDSPILQEAFOM-----VIAEKDLKNIRPADGSDYP	335						
Db	442	kqkaaea---tkvaeekqkaaeatkvaeeekqkaaeatkvaeeekq---kaaeatkva	494						
QY	336	NPGETTVGSGKQGGSIGSIRVSYMLLDDAENETASILMSGFRQMIHMFNTENPDSQAAQOE	395						
Db	495	e-----aeekkaaeatkva-----aeekqkaaeatk-----vaeekqkaaeat	533						
QY	396	LAAQARAAKAAGDDSAALAAADAAOKALEAALGKAGQOQGIL	436						
Db	534	kvaeekqkaa--eatkvaaekqkaaeakamesqkgrfl	572						
RESULT	15								
AAG29572		ID	AAG29572 standard; Protein; 571 AA.						
XX		AC	AAG29572;						
XX		DT	17-OCT-2000 (first entry)						
XX		DE	Arabidopsis thaliana protein fragment SEQ ID NO: 35209.						
XX		KW	Protein identification; signal transduction pathway; metabolic pathway;						
XX		KW	hybridisation assay; genetic mapping; gene expression control; promoter;						
XX		KW	termination sequence.						
OS			Arabidopsis thaliana.						
XX		PN	EP1033405-A2.						
XX		PD	06-SEP-2000.						
XX		PF	25-FEB-2000; 2000EP-0301439.						
XX		PR	25-FEB-1999; 990S-0121825.						
PR		PR	05-MAR-1999; 990S-0123180.						
PR		PR	09-MAR-1999; 990S-0123548.						
PR		PR	23-MAR-1999; 990S-0125788.						
PR		PR	25-MAR-1999; 990S-0126264.						
PR		PR	29-MAR-1999; 990S-0126785.						
PR		PR	01-APR-1999; 990S-0127462.						
PR		PR	06-APR-1999; 990S-0128234.						
PR		PR	08-APR-1999; 990S-0128714.						
PR		PR	16-APR-1999; 990S-0129845.						
PR		PR	19-APR-1999; 990S-0130077.						
PR		PR	21-APR-1999; 990S-0130449.						
PR		PR	23-APR-1999; 990S-0130510.						
PR		PR	28-APR-1999; 990S-0130891.						
PR		PR	28-APR-1999; 990S-0131449.						
PR		PR	30-APR-1999; 990S-0132048.						











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:36:13 ; Search time 85.91 seconds  
(without alignments)  
128.351 Million cell updates/sec

Title: US-09-391-606-8  
Perfect score: 2412  
Sequence: 1 MNP1GPGIDETERTPPAD.....QKLISEDLNSAVDHHHHH 490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152	6.3	1561	3	US-08-894-017-23
2	150.5	6.2	643	2	US-08-216-894-8
3	150.5	6.2	643	4	US-09-115-746-8
4	147	6.1	1566	2	US-08-687-956A-23
5	146.5	6.1	1565	6	5352450-2
6	138.5	5.7	564	2	US-08-216-894-2
7	138.5	5.7	564	4	US-09-115-746-2
8	138	5.7	2842	1	US-07-741-940-7
9	138	5.7	2842	1	US-08-289-548A-7
10	138	5.7	2842	1	US-08-452-654-7
11	138	5.7	2843	1	US-07-741-940-2
12	138	5.7	2843	1	US-08-289-548A-2
13	138	5.7	2843	1	US-08-452-654-2
14	138	5.7	2843	1	US-08-452-655B-2
15	138	5.7	2843	1	US-08-452-655B-7
16	138	5.7	2843	2	US-08-370-235A-2
17	138	5.7	2843	3	US-08-450-582-2
18	138	5.7	2843	3	US-08-450-582-7
19	138	5.7	2973	2	US-08-821-355A-7
20	138	5.7	2973	2	US-09-003-687A-7
21	138	5.7	2973	4	US-09-136-605-7
22	136.5	5.7	1154	4	US-08-974-549A-611
23	136.5	5.7	2285	4	US-09-308-375-2
24	135.5	5.6	639	1	US-08-466-390-2
25	135.5	5.6	639	1	US-08-470-950-2
26	135.5	5.6	639	1	US-08-467-781-2
27	135.5	5.6	639	1	US-08-195-487-2

28	135.5	5.6	639	2	US-08-483-924-2	Sequence 2, Appli
29	135.5	5.6	639	5	PCT-US93-06160-2	Sequence 2, Appli
30	135	5.6	885	2	US-08-533-306A-4	Sequence 4, Appli
31	135	5.6	885	2	US-08-742-923A-4	Sequence 4, Appli
32	131.5	5.5	593	2	US-08-591-079-8	Sequence 8, Appli
33	131.5	5.5	593	2	US-08-591-079-10	Sequence 10, Appli
34	130	5.4	2035	1	US-08-046-585-5	Sequence 5, Appli
35	130	5.4	2035	1	US-08-393-703-5	Sequence 5, Appli
36	130	5.4	2035	5	PCT-US93-11721-5	Sequence 5, Appli
37	129.5	5.4	816	2	US-08-533-306A-6	Sequence 6, Appli
38	129.5	5.4	816	2	US-08-742-923A-6	Sequence 6, Appli
39	127.5	5.3	635	4	US-08-931-608A-5	Sequence 5, Appli
40	126.5	5.2	1093	5	PCT-US93-03077-1	Sequence 1, Appli
41	126	5.2	466	4	US-08-235-836C-107	Sequence 107, App
42	125	5.2	2101	1	US-08-466-390-4	Sequence 4, Appli
43	125	5.2	2101	1	US-08-470-950-4	Sequence 4, Appli
44	125	5.2	2101	1	US-08-467-781-2	Sequence 4, Appli
45	125	5.2	2101	1	US-08-195-487-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-08-894-017-23  
; Sequence 23, Application US/08894017  
; Patent No. 6024958  
; GENERAL INFORMATION:  
; APPLICANT: Lehner, Thomas  
; APPLICANT: Kelly, Charles  
; TITLE OF INVENTION: POLYPEPTIDE FRAGMENTS CAPABLE  
; TITLE OF INVENTION: OF COMPETITION WITH STREPTOCOCCUS MUTANS ANTIGEN I/II  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/894,017  
; FILING DATE: 31-JUL-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/00207  
; FILING DATE: 31-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murshige, Kate H  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 25150-20067.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-887-1500  
; TELEFAX: 202-822-0168  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1561 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-894-017-23

Query Match 6.3%; Score 152; DB 3; Length 1561;  
Best Local Similarity 22.3%; Pred. No. 0.004;  
Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;



```

Qy 48 KESTDTSVERWSILRSVAVNALMSLADKLGITASSN-SSSTSRSDVDSTTATPTPPPT 106
Db 2 KVKTYGFRKSKISKTLCGAVLGTVAAVSVAGQKVFADETTTSDVTKVVGQTGNPAT 61
Qy 107 -----SDDYKTKQATAYDTIFTST-----SLADIAALYS-LQDAVNIKDT 147
Db 62 NLPEAGSASQAQSQSKLERQMVHTIEVPKTDQAADKAKSAGVNVQDADVN-KGT 120
Qy 148 AATDEETAIAAEWE-----TKNADAIAK-----YGAQITELAKYASNDQAILSLGK 193
Db 121 VKTAE-AYQRETEIKEDYTKQAEIKKTDQYKSDVAAHEAEVAKIKAKNOATKEQYK 179
Qy 194 LTFDILL-OTALLQSVANNKAAEL-----LKEMQDNPVPGKPTATAQSLVDQTDATAT 247
Db 180 ----DMVAHKAEVERINAANAATAYEAKLAQYADLAQAAVOKTNAANQASY-OKALAA 234
Qy 248 OIE-----KDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAK-----TQIAEA 299
Db 235 QAEKRVQEAANAAKAAV---DTAVANNKNTETAAANEELIRKRNATAKAEYETKLAQY 291
Qy 300 Q-----KKFPDPSPILOAE-OMVIOA-EKDLNKKPAD-----GSDVNPNGTGVG 342
Db 292 QAEKRVQEAANAADYQAKLTAYOTELARVQKANADAKAAAYEAAVANAANKNAALTAE 351
Qy 343 GS-----KQGSIGSIRVSMILLDDAENETASILMSGPFROMIHMENTE 385
Db 352 NTAIKORNENAKATYEAALQYEDLAQVAAVKKANAANEADYQAK---LTAYOTELARVQKA 408
Qy 386 NPDQAAQOELAAQARAQAAK--GDSAAAAL-ADAQKALEAALGK 428
Db 409 NADAKAAYEAAVANAANAALTAENTAIKKRNADAKADYEAKLAK 454

```

RESULT 2

```

US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-8

```

```

Query Match 6.2%; Score 150.5; DB 2; Length 643;
; Best Local Similarity 21.5%; Pred. No. 0.0015;
; Matches 99; Conservative 58; Mismatches 169; Indels 135; Gaps 17;
Qy 7 PGPIDETERTPPADLSAAGLEASAAKSAEQAQIAGAEAKPKESKTSVERWSILRSVAVN 66
Db 216 PPKSDDLPGSPSQL-QQAENNTSKKEMTKLRKVKAEEKLDINRAIKLEEERN 273
Qy 67 ALMSLADKLGITASSNSSSTSRSDVDSTTATPTPPPTSDDYKTKQATAYDTIFTSTS 126
Db 274 QAYKAAHK-----ABEEKAKTQRLITTESENINLKKRP----- 307
Qy 127 LADIAALVSLQDAVN-----IKDTAATDE-----ETAIAAEWETKNADAIAKVGAI 174
Db 308 -----NDVSNRDKKKNSETAKTDEVEKQRAAEAAKAVETEKQRA-----AEA 350
Qy 175 TELAKYASNDQAIILDSLGKLTSPDLLQTLALQSVANNKAAELLKEMQDNPVPGKTPAI 234
Db 351 TKVAE-----AEKRAAEAAKAVETEKQRAAEATKV 381
Qy 235 A-----QSLVDQTDATATQIEKDNAGIDAYFAGQNASGAVENAKSN-----NSISNID 283
Db 382 AEAQKQAAEAAKAVETEKQRAAEATKVAAEAKQRAAEAMKVAAEAKQAAEATKVAAEAE 441
Qy 284 SAKAAIATAKTOIAEAK-KFPDPSPILOEAEQM-----VIOAEKDLNKKPADGSDVP 335
Db 442 KQKAAE---TKVAAEAKQAAEATKVAAEAKQKAAEATKVAAEAKQ-----KAAEATKVA 494
Qy 336 NPGTGVGSKQSGSIGSIRVSMILLDDAENETASILMSGPFROMIHMENTENPDSQAQOE 395
Db 495 E-----AEKQAAEATKVA-----EAEKQAAEATK-----VAAEAKQKAAEAT 533
Qy 396 LAQARAARAAAGDDSAALADAKQALEAALGKAGQOQGL 436
Db 534 KVAEAEKQKAA--EATKVAEAEKQKAAEAAKAAKAMESQKORFL 572

```

RESULT 3

```

US-09-115-746-8
; Sequence 8, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399

```



TELEX: 904136  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-115-746-8

Query Match 6.2%; Score 150.5; DB 4; Length 643;  
Best Local Similarity 21.5%; Pred. No. 0.0015;  
Matches 99; Conservative 58; Mismatches 169; Indels 135; Gaps 17;

QY 7 PGIDETPPADLSAAGLEASAAKSAEAOIAGAEAPKESKTSVERWSTLRSVAVN 66  
DB 216 PPKSDLVPRGSPSL--QOENNTITNSKKEMTKLREKVKKAEKELDAINRATKLEERN 273  
QY 67 ALMSLADKLGIASSSSSTSRSDVDSTTATPTPPPTSDDYKTOAQATYDTFTSTS 126  
DB 274 QAYKAAHK--AEBEKARTFORLITFESINLKKRP----- 307  
QY 127 LADTQAAVLSDAVTN-----IKDTAATDE-----ETAIAAEWETKNADAIVGAQI 174  
DB 308 -----NDAVSNRDKKNSETAKTDEVEKORAAEAAKAVETEKORA-----AEA 350  
QY 175 TELAKYASDNQAILDLSGLKLTSPDLLQTLQSVANNKKAELKEMODNPVPGKTPAI 234  
DB 351 TKVAE-----AEKKAEEAAKAVETEKORAAEATKV 381  
QY 235 A-----QSLVDQTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSN-----NSISNID 283  
DB 382 AEAEKQKAAEAKAVETEKORAAEATKVAAEAKQRAEAMKVAEAEKQAAEATKVAAEAE 441  
QY 284 SAKAAIATATQIAEAK-KFPDSPILQEAQM-----VIOAEKDLKNKPKADGSDVP 335  
DB 442 KQKAAEA---TKVAEAEKQAAEATKVAAEAEKQKAAEATKVAAEAEKQ---KAAEATKVA 494  
QY 336 NPGTGVGSKQGGSSIGSIRVSMLLDDAENETASILMSGFQIMHFNENTENPDSDQAQOE 395  
DB 495 E-----AEKQKAAEATKV-----EAQKQAAEATK-----VAEAEKQKAAEAT 533  
QY 396 LAAQARAAGDSDSAAALADAKALEALGKAGQGGIL 436  
DB 534 KVAEAEKQKAA--EATKVAAEAEKQKAAEAAKAAKAMESQKQREL 572

RESULT 4  
US-08-687-956A-23  
Sequence 23, Application US/08687956A  
Patent No. 5861157  
GENERAL INFORMATION:  
APPLICANT: BURNIE, JAMES P  
APPLICANT: MATTHEWS, RUTH C  
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF  
TITLE OF INVENTION: INFECTIONS DUE TO STREPTOCOCCI AND ENTEROCOCCI  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESSES:  
ADDRESS: PILLSBURY, MADISON & SUTRO, LLP  
STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,956A  
FILING DATE: 29-JUL-1996

CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9401689.6  
FILING DATE: 28-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 50885/222892  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/861-3000  
TELEFAX: 202/822-0944  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1566 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus sobrinus  
US-08-687-955A-23

Query Match 6.1%; Score 147; DB 2; Length 1566;  
Best Local Similarity 22.6%; Pred. No. 0.0097;  
Matches 106; Conservative 75; Mismatches 209; Indels 80; Gaps 19;  
QY 48 KESKTSVERWSTLRSVAVNALSADKLGIASSSSSTSS--RSADVSTTATAPTPTPP 105  
DB 10 RKSRISRTLCALLGTAL--LASVTGOKALAEETSTSTSGVNTAVVGTGTGNATNLPD 67  
QY 106 TSDDYKTOAQATY-----DFTFTSLADIAQALVSLQDAVNTKIDTAATDEATAA 158  
DB 68 KQDNPSSQAEATSOAAGOKTGAMSDVSVSTSELDEAAKSAQAEAGVTVSQDATVDKGTVETS 127  
QY 159 EWEETKNADAIK--VGAOITELAKYASD-----NOATLDSGLKLTSPDLLQTLQSV 209  
DB 128 DEANQKETEIKDDYSKQAADIOKTTEDYKAAVARNQAEETD---RITQENAAKKAQYEODL 184  
QY 210 NNKAAELLKEMODNPVPGKTPAIAOSLSVDOTDATATQIEKDGNAIGDAYFAGQNASGA 269  
DB 185 AARK-AEVERITNEN-----AQRKADY-EAKLAQYOKDLAAVOAANDSOAAVAA 232  
QY 270 VERAKSNNSISNIDSAKAAIATAKTOIAEA-----OKKFPDSPILQ-----EAE 313  
DB 233 AKEA-YDKELARVOAANAA---AKREYEELAAANTTNEQIKAEENATQOORNAQAKADYE 288  
QY 314 QMVIQAEKDLKNKPKADGSDVPNPGTTVGGSKQGGSSIGSIRVSMLLDDAENETASILMS 373  
DB 289 AKLAQYKDLAAAOAGNATNEADYQAKKAAVEOELARV-----QAANAANK---Q 335  
QY 374 GFQMIHMFNTENPDSDQAQOELAAQAPAKAAGDDSSAAALADAKALEALGKAGQGG 433  
DB 336 AYEQAALAAANTAKN--AQITAEENAIQORNAQAKA--NYEAKLAQYQKDLAAA-----QS 385  
QY 434 GILNALGOIASAAVVSAGVLPLOQVLWIRARYQAYVEOKLISEEDLNSAV 483  
DB 386 G--NAANEADYQEKLAAYEKELARVQAAANAANKOEYKOEKQVQANAKNAEI 433  
RESULT 5  
5352450-2  
Patent No. 5352450  
APPLICANT: KOGA, TOSHIHIKO; OKAHASHI, NOBUO; TAKAHASHI, ICHIRO;  
SHIBUYA, KOJI; OHTA, HIROTAKA  
TITLE OF INVENTION: METHOD FOR PREPARING VACCINE FOR DENTAL  
CARRIES AND VACCINAL COMPOSITIONS FOR DENTAL CARRIES USED AS NASAL  
DROP  
NUMBER OF SEQUENCES: 2  
CURRENT APPLICATION DATA:



APPLICATION NUMBER: US/07/529,602  
 FILING DATE: 29-MAY-1990  
 SEQ ID NO: 2  
 LENGTH: 1565  
 5352450-2

Query Match 6.1%; Score 146.5; DB 6; Length 1565;  
 Best Local Similarity 21.8%; Pred. No. 0.01;  
 Matches 101; Conservative 81; Mismatches 190; Indels 91; Gaps 21;

48 KESKTSVERWSILRSVAVNALMSLADKLGIASSN--SSSTSRSDVDSTATATPPPPP 105  
 Db 2 KVKATYGRKSKISKLCGAVTAAVAVAGQKVFADETTTSDVDFVGVGTQGNPA 61  
 QY 106 T-----SDYKTOAQATYDTFTST-----SLADIQAALVS-LQDAVNTIKD 146  
 Db 62 TQLEAQSASKEAEQSTKLERQMVHTIEVPKTDLDQAADAKSAGVNVVQADVN-KG 120  
 QY 147 TAATDETAIAEWE-----TKNDAIK-----VGAQITELAKYASDNQAILDSLG 192  
 Db 121 TVKTPPE-AVGKTEIKEDYTKQAEIKKTTDQTKSDVAHEAEVAKIKAKNQATKEQYE 179  
 QY 193 KLTSPDLLQALLOSANNKAA--ELLKEMODNPVPGKTPAQAQSLVDQTDATATQIE 250  
 Db 180 KMAAHKAEVERINANAASKTAYEAKLAQYQADLAQVQKTNANQAAY-QKALAAVQAE 238  
 QY 251 ----KGNAGIDAYFAGNAGAVENAKSNNSISNIDSAAKATATAK-----TQIAEAQ-- 300  
 Db 239 LKRVQEAANAANKAA--DTAANAANKTEIAAANAEIRKRNATAKAAYETKLAQYQAE 295  
 QY 301 -KKPDPSPILQAE-QWVIOA-EKDLNIRPAD-----GSDVNPVGTGGS- 344  
 Db 296 LKRVQEAANAANEADYQAKLTAYQELARVQKANADAKATYEAANAANKNAALTAENTA 355  
 QY 345 -----KQSGSIGSVIRVSMILLDDAENETASILSGFRQIMHMENTNP 388  
 Db 356 IKERNENAKATYEAALQYQADLAQVQKANAANEADYQAK-----LTAYQTELARVQKANAD 412  
 QY 389 SOAQOELAAQARAQAA--GDDSAAL-ADAQKALEAALGK 428  
 Db 413 AKAAEYAAVANAANAALTAENTAIRKRNADAKADYEAELK 455

RESULT 6  
 US-08-216-894-2  
 ; Sequence 2, Application US/08216894  
 ; Patent No. 5876734  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kirchhoff, Louis V.  
 ; APPLICANT: Otsu, Keiko  
 ; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
 ; TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington, D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/216.894  
 ; FILING DATE: 24-MAR-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 564 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-216-894-2

Query Match 5.7%; Score 138.5; DB 2; Length 564;  
 Best Local Similarity 21.4%; Pred. No. 0.01;  
 Matches 97; Conservative 54; Mismatches 165; Indels 137; Gaps 17;

QY 7 PGIDETERTPADLSAQGLEASAANKSAEAGRIAGAEAKPKESKTDSDVERWSILRSVAVN 66  
 Db 216 PKKSDLVPRGSPSQL--QQAENNITNSKKEMTKLREKVKKAKEKLDAINRATKLEERN 273  
 QY 67 ALMSLADKLGIASSNSSSTSRSDVDSTATATPPPTSDDYKTOAQATYDTFTST 126  
 Db 274 QAYKAAHK---AEEERAKTFQRLITFESINILKKRP----- 307  
 QY 127 LADIQAALVSLQDAVTN-----IKDTAATDE-----ETAIAAEWETKNADAIVGQAI 174  
 Db 308 -----NDAVSNROKKKNSSETAKTDEVEKQRAAEAAKAVETEKQRA-----AEA 350  
 QY 175 TELAKYASDNQAILDSGLKLTSPDLLQALLOSANNKAAELLKEMODNPVPGKTPAI 234  
 Db 351 TKVAE-----AEKRAAEAAKAVETEKQRAAEATKV 381  
 QY 235 A---OSLVDOTDATATQIEKDGNAIGDAYFAGNAGAVENAKSNNSISNIDSAAKAA 290  
 Db 382 AEAEKQAAEAAKAVETEKQRAAEATKVAEAEKQRAEAMK-----VAEAEKQAAEA 434  
 QY 291 TAKTOIAEAKQ-KFPDPSILQAEQM-----VIOAKDLKNIKPADSGSDVNPVGTG 342  
 Db 435 ---TKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQ-----KAAEATKVAE----- 481  
 QY 343 GSKOOGSSIGSVIRVSMILLDDAENETASILSGFRQIMHMENTNPDSQAAQOELAAQARA 402  
 Db 482 AEKQAAEATKVA-----EAERKQAAEATK-----VAEAEKQAAEATKVAEAEK 526  
 QY 403 AKAAAGDDSAAAA-----LADAQKALEAALGKAGQ 431  
 Db 527 QKAAEATKVAEAEKQAAEATKVAEAEKQKAGE 559

RESULT 7  
 US-09-115-746-2  
 ; Sequence 2, Application US/09115746  
 ; Patent No. 6228601  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kirchhoff, Louis V.  
 ; APPLICANT: Otsu, Keiko  
 ; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
 ; TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington, D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/115,746  
 ; FILING DATE:



```

QY 11 DETERPTPADLSA-QGLEASAAANKSAEQAORIAEAKPE-----SKTDSV----- 55
Dd 1658 DLJIESPPNELAAGEGVRGAQSGFEKRDPTIETGRSTDEAAGKTSVVTIPELDONKA 1717
QY 56 ERSSILRSVNALM-----SLADKLGIASSNSSSTSRSDVDSTTATPTPP 103
Dd 1718 EEGDILAEICINSAMPKSKHKPRVKKIMDQVQASASSASPKNQLDGGKKKPTSPVKP 1777
QY 104 PPTSDDYKTAQATAYDT-----IFT-----STSLADLQAALVSLQDAV--- 141
Dd 1778 IPONTEYRVRKNADSKKNLNAERVFSNDKSKONLKNKSKOFNDKLPNNEORVRGSF 1837
QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIKVGAQITELAKYASDQAILDSL 191
Dd 1838 AFDSPHHYPIEGTPYCFSRNDSLSDPDDVDLSREKAEL---RKAKEN---ESE 1890
QY 192 GKUTSSEDLLOTALLOSVAANNKAELLKPM-----ODNPVVPKGTPATAIOSLVODTATAT 247
Dd 1891 AKYTS-----HTELTSNOQSANTQAIAKQPINRGOPKPILO-KOSTFPQSSKDIPDRGAA 1945
QY 248 QIEKDGNAGIDAVFAGONASGAVENAKSNNSISNIDSAKAALATAKTQIAEAOKKFPDSP 307
Dd 1946 TDEKLQN-----FALNETPVCFSHNSLSLSLSDID----- 1975
QY 308 ILQEAQOMVIAQEKDLKNLKPADGSDVPNPGTTVGSGKO-----QGSISGISIR 355
Dd 1976 --QENNN---KENEPKITEPTDQSGEPSKPOASGYAPKSFHVEDTPVCFSRNSSLSSLS 2030
QY 356 VSMLLDDAENETASILMSGFROMIHFNFTENPDSQAQOELAAQAAAKAAGDSDSAAAL 415

```



Db 2031 I-----DSEDDLLQECISS-----AMPKPKKPSRLKGDNEKHSRNMGGILGED-LTLDL 2079

QY 416 ADAOKA-LEAALG-----KAGQQ--QGILNALGQIASAAVVS 449  
 Db 2080 KDQRPDSEHGLSPDSENFDMKALQEGANSIVSSHLHQAAAACLS 2124

RESULT 9

US-08-289-548A-7  
 ; Sequence 7, Application US/08289548A  
 ; Patent No. 5648212

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS  
 APPLICANT: ANAND, RAKESH  
 APPLICANT: CARLSON, MARY  
 APPLICANT: GRODEN, JOANNA  
 APPLICANT: HEDGE, PHILIP J.  
 APPLICANT: JOSLYN, GEOFF  
 APPLICANT: KINZLER, KENNETH  
 APPLICANT: MARKHAM, ALEXANDER F.  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Allegretti, LTD  
 STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/289,548A  
 FILING DATE: 12-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.  
 REGISTRATION NUMBER: 32,141  
 REFERENCE/DOCKET NUMBER: 1107.46943  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-508-9100  
 TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2842 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: APC

US-08-289-548A-7

Query Match 5.7%; Score 138; DB 1; Length 2842;

Best Local Similarity 20.0%; Pred. No. 0.11;

Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERTPPADLSA-OGLESAANKSAEORAGAEAKPE-----SKTDSV-----55

Db 1658 DLTIESPNNELAAGEVGRGAQSGEPEKRDITPTIGRSTDEAOGGKTSSVTIPELDDNKA 1717

QY 56 ERWSILERSAVNALM-----SLADKLGIASSNSSSTSRSDVDSITATPTPP 103

Db 1718 EEGDILAEINCINSAMPKGSHPKPRVKKIMDQVOQAASSASPKNKQLDGGKKKPTSPVKP 1777

QY 104 PPTSDYKTKAQATYDT-----IFT-----STSLADIAQALVLSQDAV----141  
 Db 1778 IPONTEYTRVRKNADSKNNLNAERVFSNKNKSKONLNKNSKDFNCKLPNNEDRVGSGF 1837  
 QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDSL 191  
 Db 1838 AFDSPPHYTPIEGTPYCFSRNDSLSLDFDDDDVDLSREKAEL-----RKAENK---ESE 1890  
 QY 192 GKLTSEDLLOTALLOSANNKAAELLEM-----QNPVVVPGKTPAIAOSLVDTOTATAT 247  
 Db 1891 AKVTS-----HTELTSNQOSANKTOAIAKQPINRGPKPIQ-KOSTFPQSSKDIPDRGAA 1945  
 QY 248 QIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSARAAIATATQTAIAEAKKFPDSP 307  
 Db 1946 TDEKLQN-----FAIENTPVCFSHNSSLSLSDID-----1975  
 QY 308 ILQEAQMVIQAEKDLKNIKPADGSDVPNPGTTVGSGKQ-----QGSSIGSIR 355  
 Db 1976 --QENNN--KENEPKETEPDPSQGEPSKPQASGVAPKSFHVEDTPVCFSRNSSLSSLS 2030  
 QY 356 VSMLLDDAENETASILMSGFRMIHMENTENPDQSQAAQOELAAQAAKAAAGDSDSAAAL 415  
 Db 2031 I-----DSEDDLLQECISS-----AMPKPKKPSRLKGDNEKHSRNMGGILGED-LTLDL 2079  
 QY 416 ADAOKA-LEAALG-----KAGQQ--QGILNALGQIASAAVVS 449  
 Db 2080 KDQRPDSEHGLSPDSENFDMKALQEGANSIVSSHLHQAAAACLS 2124

RESULT 10

US-08-452-654-7

; Sequence 7, Application US/08452654  
 ; Patent No. 5691454

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS  
 APPLICANT: ANAND, RAKESH  
 APPLICANT: CARLSON, MARY  
 APPLICANT: GRODEN, JOANNA  
 APPLICANT: HEDGE, PHILIP J.  
 APPLICANT: JOSLYN, GEOFF  
 APPLICANT: KINZLER, KENNETH  
 APPLICANT: MARKHAM, ALEXANDER F.  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: THLIVERIS, ANDREW  
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner, Birch, McKie & Beckett  
 STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/452,654  
 FILING DATE: 25-MAY-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/741,940

FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.035574

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100



TELEFAX: 202-508-9299  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2842 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: APC  
 US-08-452-654-7

Query Match 5.7%; Score 138; DB 1; Length 2842;  
 Best Local Similarity 20.0%; Pred. No. 0.11; Indels 144; Gaps 23;  
 Matches 105; Conservative 82; Mismatches 194;

Qy 11 DETERTPPADLSA-QGLEASAAKSAEQAQRIAGAEAKPKE-----SKTDSV----- 55  
 Db 1658 DLTIESPPELAAGBVGGAQSGFEKRDITPTGRSTDEAOGKTSSTVIPELDDNKA 1717  
 Qy 56 ERWSILRSVAVNLM-----SLADKLGIASSNSSSTSRSDVSTTATATPTPP 103  
 Db 1718 EEGDILAEICINSAMPKSHKPFVRVKKIMDQVQOASASSAPNKLQDGGKKKPTSPVKP 1777  
 Qy 104 PPTSDDYKTOAQATYDT-----IFT-----STSLADIQAALVSLQDAV----- 141  
 Db 1778 IPQNTFYTRVRKNADSKNNLNAERVFSDNKKQKNNKSKDFNDKLPNNEDRVGRSF 1837  
 Qy 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNOAILDSL 191  
 Db 1838 AFDSPHHVTPTEGTPYCFSRNDSLSSLDLDDDDVDSREKAEL-----RKAKENK---ESE 1890  
 Qy 192 GKLTISFDLLQTLLOSVAANNKAAELLKEM-----QDNVPVPGKTPAIAQSLVDQTDATAT 247  
 Db 1891 AKVTS-----HTELTSNQQSANKTOAIAKQPINRGQPKPIQ-KQSTFPQSSKDIPDRGAA 1945  
 Qy 248 QLEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAIATAKTQIAEAKKFPDPS 307  
 Db 1946 TDEKLN-----FAIENTPVCFSHNSLSLSDID----- 1975  
 Qy 308 ILQEAQMVIOAEKDLNKPADGSDVPNPGTTGGSKO-----QGSSIGSIR 355  
 Db 1976 --QENNN---KENEPKETEPDQSGEPKPAQSGAPKSFHVEDTPVCFSRNSSLSSLS 2030  
 Qy 356 VSMLLDDAENETASILMSGFQMTMFNTENPDQSQAAQOELAAQARAAGDDSAAL 415  
 Db 2031 I-----DSEDDLQECISS-----AMPKKKPSRLKGNKHSRPNMGGIIGED-LTLDL 2079  
 Qy 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQIASAAVVS 449  
 Db 2080 KDIQRPDSEHGLSPDSENFDMKATQEGANSIVSSLHQAAAAACLS 2124

RESULT 11  
 US-07-741-940-2  
 Sequence 2, Application US/07741940  
 Patent No. 5352775  
 GENERAL INFORMATION:  
 APPLICANT: ALBERTSEN, HANS  
 APPLICANT: ANAND, RAKESH  
 APPLICANT: CARLSON, MARY  
 APPLICANT: GRODEN, JOANNA  
 APPLICANT: HEDGE, PHILIP J.  
 APPLICANT: JOSLYN, GEOFF  
 APPLICANT: KINZLER, KENNETH  
 APPLICANT: MARKHAM, ALEXANDER F.  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: THLIVERIS, ANDREW  
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 94  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Banner, Birch, McKie & Beckett  
 STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20001-4598  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07741,940  
 FILING DATE: 19920109  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kagan, Sarah A.  
 REGISTRATION NUMBER: 32,141  
 REFERENCE/DOCKET NUMBER: 1107.035574  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-508-9100  
 TELEFAX: 202-508-9299  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2843 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-741-940-2

Query Match 5.7%; Score 138; DB 1; Length 2843;  
 Best Local Similarity 20.0%; Pred. No. 0.11; Indels 144; Gaps 23;  
 Matches 105; Conservative 82; Mismatches 194;  
 Qy 11 DETERTPPADLSA-QGLEASAAKSAEQAQRIAGAEAKPKE-----SKTDSV----- 55  
 Db 1659 DLTIESPPELAAGBVGGAQSGFEKRDITPTGRSTDEAOGKTSSTVIPELDDNKA 1718  
 Qy 56 ERWSILRSVAVNLM-----SLADKLGIASSNSSSTSRSDVSTTATATPTPP 103  
 Db 1719 EEGDILAEICINSAMPKSHKPFVRVKKIMDQVQOASASSAPNKLQDGGKKKPTSPVKP 1778  
 Qy 104 PPTSDDYKTOAQATYDT-----IFT-----STSLADIQAALVSLQDAV----- 141  
 Db 1779 IPQNTFYTRVRKNADSKNNLNAERVFSDNKKQKNNKSKDFNDKLPNNEDRVGRSF 1838  
 Qy 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNOAILDSL 191  
 Db 1839 AFDSPHHVTPTEGTPYCFSRNDSLSSLDLDDDDVDSREKAEL-----RKAKENK---ESE 1891  
 Qy 192 GKLTISFDLLQTLLOSVAANNKAAELLKEM-----QDNVPVPGKTPAIAQSLVDQTDATAT 247  
 Db 1892 AKVTS-----HTELTSNQQSANKTOAIAKQPINRGQPKPIQ-KQSTFPQSSKDIPDRGAA 1946  
 Qy 248 QLEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAIATAKTQIAEAKKFPDPS 307  
 Db 1947 TDEKLN-----FAIENTPVCFSHNSLSLSDID----- 1976  
 Qy 308 ILQEAQMVIOAEKDLNKPADGSDVPNPGTTGGSKO-----QGSSIGSIR 355  
 Db 1977 --QENNN---KENEPKETEPDQSGEPKPAQSGAPKSFHVEDTPVCFSRNSSLSSLS 2031  
 Qy 356 VSMLLDDAENETASILMSGFQMTMFNTENPDQSQAAQOELAAQARAAGDDSAAL 415  
 Db 2032 I-----DSEDDLQECISS-----AMPKKKPSRLKGNKHSRPNMGGIIGED-LTLDL 2080  
 Qy 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQIASAAVVS 449  
 Db 2081 KDIQRPDSEHGLSPDSENFDMKATQEGANSIVSSLHQAAAAACLS 2125



Query Match



Best Local Similarity 20.08; Pred. No. 0.11;  
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

Qy 11 DETERTPADLSA-QGLEASANKSAEORITAGAEAKPE-----SKTDSV----- 55  
Db 1659 DLTIESPNEIAAGEVGRGAQSGEFKRDITPTEGRTDEAOGGKTSVTIPELDNKA 1718  
Qy 56 ERWSTLSRAVNAIM-----SLADKLGIASSNSSTSRSDVDSTTATPTPP 103  
Db 1719 EGGDILACINSAMPKSHKHPFRVKKIMDVOQASASSAPNKNQDGLPNNEDVRGSGF 1778  
Qy 104 PPTSDYKTOQATYDT-----IFT-----STLADIQAALVSLQDAV----- 141  
Db 1779 IPQNTYRTRVRKNADSKNNLNAERVFSDNKKNSKDFNDKLPNNEDVRGSGF 1838  
Qy 142 -----TNIKDT--AATDETAIAAEWETKNADAIVGAQITELAKYASDNQAILDSL 191  
Db 1839 AFDSPHHYTPTEGTPYCFSRNDSLSLDFDDDDVLSREKAE-----RKAKENK---ESE 1891  
Qy 192 GKLTSDLLQALLOSVAANNKAELLKEM-----QDNVPVPGKTPAIAQSLVDQTDATAT 247  
Db 1892 AKVTS-----HTELTSNQOSANKTOAIAKOPINRGOPKPILO-KOSTFFQSSKDIPDRGAA 1946  
Qy 248 QIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAAATATAKTQIAEAKKFPDPS 307  
Db 1947 TDEKLQN-----FAIENTPVCFSHNSLSLSLSDID----- 1976  
Qy 308 ILQEAQMVIOAEKDKLNKIPADGSDVPNPGTTVCGSKQ-----QGSSIGSIR 355  
Db 1977 --QENNN--KENEPKETEPPDSOGEPKQASGYAPKSPHVEDTPVCFSRNLSLSLS 2031  
Qy 356 VSMILDDAENETASTILMSGFRQIMHMTENPDSQAQOELAAQARAAGDSDSAAAL 415  
Db 2032 I-----DSEDDILQBCISS-----AMPKKKPSRLKGNKHSRPNMGMGILGED-LTLDL 2080  
Qy 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQTASAAVVS 449  
Db 2081 KDQIRPDSEHGLSPDSENFDMKATQEGANSIVSSLHQAAAAACLS 2125

RESULT 14  
US-08-452-655B-2  
; Sequence 2, Application US/08452655B  
; Patent No. 5783666  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
; NUMBER OF SEQUENCES: 102  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Wilcoff, Ltd.  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/452,655B  
; FILING DATE: 25-MAY-1995

CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,548  
; FILING DATE: 12-AUG-1994  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/741,940  
; FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-452-655B-2

Query Match 5.78; Score 138; DB 1; Length 2843;  
Best Local Similarity 20.08; Pred. No. 0.11;  
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

Qy 11 DETERTPADLSA-QGLEASANKSAEORITAGAEAKPE-----SKTDSV----- 55  
Db 1659 DLTIESPNEIAAGEVGRGAQSGEFKRDITPTEGRTDEAOGGKTSVTIPELDNKA 1718  
Qy 56 ERWSTLSRAVNAIM-----SLADKLGIASSNSSTSRSDVDSTTATPTPP 103  
Db 1719 EGGDILACINSAMPKSHKHPFRVKKIMDVOQASASSAPNKNQDGLKKTSPVKP 1778  
Qy 104 PPTSDYKTOQATYDT-----IFT-----STLADIQAALVSLQDAV----- 141  
Db 1779 IPQNTYRTRVRKNADSKNNLNAERVFSDNKKNSKDFNDKLPNNEDVRGSGF 1838  
Qy 142 -----TNIKDT--AATDETAIAAEWETKNADAIVGAQITELAKYASDNQAILDSL 191  
Db 1839 AFDSPHHYTPTEGTPYCFSRNDSLSLDFDDDDVLSREKAE-----RKAKENK---ESE 1891  
Qy 192 GKLTSDLLQALLOSVAANNKAELLKEM-----QDNVPVPGKTPAIAQSLVDQTDATAT 247  
Db 1892 AKVTS-----HTELTSNQOSANKTOAIAKOPINRGOPKPILO-KOSTFFQSSKDIPDRGAA 1946  
Qy 248 QIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAAATATAKTQIAEAKKFPDPS 307  
Db 1947 TDEKLQN-----FAIENTPVCFSHNSLSLSLSDID----- 1976  
Qy 308 ILQEAQMVIOAEKDKLNKIPADGSDVPNPGTTVCGSKQ-----QGSSIGSIR 355  
Db 1977 --QENNN--KENEPKETEPPDSOGEPKQASGYAPKSPHVEDTPVCFSRNLSLSLS 2031  
Qy 356 VSMILDDAENETASTILMSGFRQIMHMTENPDSQAQOELAAQARAAGDSDSAAAL 415  
Db 2032 I-----DSEDDILQBCISS-----AMPKKKPSRLKGNKHSRPNMGMGILGED-LTLDL 2080  
Qy 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQTASAAVVS 449  
Db 2081 KDQIRPDSEHGLSPDSENFDMKATQEGANSIVSSLHQAAAAACLS 2125

RESULT 15  
US-08-452-655B-7  
; Sequence 7, Application US/08452655B  
; Patent No. 5783666  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA



APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,655B  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107,49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
US-08-452-655B-7

Query Match 5.7%; Score 138; DB 1; Length 2843;  
Best Local Similarity 20.0%; Pred. No. 0.11;  
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;  
Qy 11 DETERTPPADLSA-OGLEPASAANKSAEAGRIAGAPKPE-----SKTDSV----- 55  
Db 1659 DLTIESPPNELAAGEVGRGAQGEFEKRDPTPTGRSTDEAQQGKTSSTVPIPELDDNKA 1718  
Qy 56 ERWSILRSVAVNALM-----SLADKLGIASSNSSSTSRSDVDSTTATPTTP 103  
Db 1719 EEGDILAEICINSAMPKSGKSHKPRVRYKIMDQVQOASASSAPNKNQLDGGKKKPTSPVKP 1778  
Qy 104 PPTSDYKTAQTAYDT-----IFT-----STSLADIQAALVSLQDAV----- 141  
Db 1779 IPQNTYETRYVRKNADSNLNAERVFSNDKSKKQNLKNSKDFNDKLPNNEDVRGGSF 1838  
Qy 142 -----TNKDT--AATDEETAIAEWETKNADAIVKGAQITELAKYASDNQAILDSL 191  
Db 1839 AFDSPHHYPTIEGTPYCFPSRNDLSLSDFDVDDVLSREKAEL----RKAKENK----ESE 1891  
Qy 192 GKLTISFDLLQTLQSVANNKAAELKEM----QDNVPVPGKTPAIAQSLVDQTDATAT 247  
Db 1892 AKVTS-----HTELTSNQSANKTAIAKQPINRGOPKPILO-KOSTFPQSKDIPDRGAA 1946

Qy 248 QLEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAAAIATAKTAQIAEAQKFPDSP 307  
Db 1947 TDEKLQN-----FAIENTPVCFSHNSLSLSLSDID----- 1976  
Qy 308 ILQAEQWVIOAEKDLKNIKPADGSDVNPCTTVGGSKQ-----QGSSIGSIR 355  
Db 1977 --QENNN--KENEPKETEPPDSQGEPSKPOASGYAPKSPHVEDTPVCFSRNSLSLS 2031  
Qy 356 VSMILLDDAENETASTILMSGFRQMTHMFTENPDSSQAQQELAAQAAKAAAGDDSAAL 415  
Db 2032 I-----DSEDDLQECISS-----AMPKKKPSRLKGNKHSRPNWGGILGED-LTLDL 2080  
Qy 416 ADAQKA-LEAALG-----KAGQQ--QGILNALQIASAIVVS 449  
Db 2081 KDIQRPDSEHGLSPDSENFDMKAIQEGANSIVSSILHQAACLS 2125

Search completed: February 7, 2002, 21:36:19  
Job time: 20488 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:38:03 ; Search time 96.2 Seconds  
(without alignments)  
387.999 Million cell updates/sec

Title: US-09-391-606-8

Perfect score: 2412

Sequence: 1 MNPICGPGPIDETERTPPAD.....QKLISEDLNSAVDHHHHHH 490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2214	91.8	715	2 I40729	hypothetical 76k p
2	2188	90.7	651	2 D72042	conserved hypothet
3	2188	90.7	651	2 E86581	CHLPN 76 kDa homol
4	451.5	18.7	647	2 G71490	hypothetical prote
5	201	8.3	973	2 C85693	probable membrane
6	190	7.9	1122	2 G84887	probable tail fibe
7	188	7.8	2055	2 T31110	extracellular matr
8	178	7.4	1822	2 S33441	EF protein - strep
9	174.5	7.2	545	2 E84327	Htr7 transducer [i
10	174.5	7.2	545	2 T46811	halobacterial tran
11	171	7.1	627	2 F84194	Htr14 transducer [
12	170	7.0	1128	2 T30296	R27-2 protein - Tr
13	166.5	6.9	2232	2 T34434	hypothetical prote
14	160.5	6.7	544	2 T44938	transducer protein
15	160.5	6.7	1156	2 T34852	probable secreted
16	158	6.6	536	1 A47190	transducer protein
17	158	6.6	536	2 E84318	Htr1 transducer [i
18	158	6.6	810	2 T46810	halobacterial tran
19	158	6.6	810	2 F84327	Htr5 transducer [i
20	158	6.6	1365	2 T30282	lmp1 protein - Myc
21	156.5	6.5	641	2 C82026	methyl-accepting c
22	156	6.5	990	2 T51618	nucleolar phosphop
23	155.5	6.4	582	2 S24545	intermediate filam
24	155	6.4	892	2 T50985	related to transcr
25	155	6.4	1147	2 T35781	hypothetical prote
26	154.5	6.4	764	2 A84328	Htr2 transducer [i
27	154.5	6.4	765	1 T44946	transducer protein
28	154.5	6.4	1262	2 T22523	hypothetical prote
29	153	6.3	642	1 T44253	transducer protein

30	152.5	6.3	860	2 T14650	tail fiber protein
31	152.5	6.3	978	2 T14968	phage lambda-relat
32	152.5	6.3	1528	2 A60338	surface antigen A
33	152	6.3	1561	1 S06839	surface antigen sp
34	152	6.3	3488	2 T34418	hypothetical prote
35	151.5	6.3	1238	2 T03465	probable exonuclea
36	150.5	6.2	643	2 H84305	Htr8 transducer [i
37	149	6.2	1566	2 A43607	cell surface antig
38	148	6.1	571	2 D86164	hypothetical prote
39	148	6.1	729	2 E70803	hypothetical prote
40	148	6.1	881	2 S56032	probable membrane
41	148	6.1	1556	2 A60988	saliva-interacting
42	147.5	6.1	1302	1 JC6009	surface-located me
43	147	6.1	778	2 T48897	transducer protein
44	147	6.1	5327	2 T13564	microtubule-associ
45	146.5	6.1	446	2 H83098	hypothetical prote

## ALIGNMENTS

### RESULT 1

I40729

Hypothetical 76k protein - Chlamydomophila pneumoniae (strain AR39)

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 31-Mar-2000

C:Accession: I40729

R:Peréz-Melgosa, M.; KUO, C.

Infect. Immun. 62, 880-886, 1994

A:Title: Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76

A:Reference number: I40729; MUID:94156481

A:Accession: I40729

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-715 <RES>

A:Cross-references: GB:L23921; NID:g435961; PIDN:AAA23117.1; PID:g435962

A:Experimental source: strain AR-39

C:Comment: This is the hypothetical translation of a sequence that was reported as t

Query Match 91.8%; Score 2214; DB 2; Length 715;

Best Local Similarity 98.9%; Pred. No. 1.6e-110;

Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY	1	MVNPICGPGPIDETERTPPADLSAQGLEASAAKSAEAOIRAGAEAKPKESKTSVERWSI	60
DB	257	LVPNPICGPGPIDETERTPPADLSAQGLEASAAKSAEAOIRAGAEAKPKESKTSVERWSI	316
QY	61	LRSAVNALMSLADKLGIASSNSSSTSRSDVDTTATATPTPPPTSDDYKTAQATAYDT	120
DB	317	LRSAVNALMSLADKLGIASSNSSSTSRSDVDTTATATPTPPPTFDDYKTAQATAYDT	376
QY	121	IFTSLSLADIQAALVSLQDAVNTIKDTATDEETAIAAEWETKNADAIVKGAQITELAKY	180
DB	377	IFTSLSLADIQAALVSLQDAVNTIKDTATDEETAIAAEWETKNADAVKVGAIQITELAKY	436
QY	181	ASDNGAILDLSLGLTFSFLLQALLOSANNKAAELKEMODNPVPGKTPFAIAQSLVD	240
DB	437	ASDNGAILDLSLGLTFSFLLQALLOSANNKAAELKEMODNPVPGKTPFAIAQSLVD	496
QY	241	QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNTSDSAKAATATAKTOIAEAQ	300
DB	497	QTDATATQIEKDGNAIRDAYFAGQNASGAVENAKSNNSISNTSDSAKAATATAKTOIAEAQ	556
QY	301	KKFPDSPILQEAQWVIOAEKDLNKPAGDSVDVNPNGTTVGSGKQSGSSIGSIRVSMLL	360
DB	557	KKFPDSPILQEAQWVIOAEKDLNKPAGDSVDVNPNGTTVGSGKQSGSSIGSIRVSMLL	616
QY	361	DDAENETASILMSGFQMIHMENTENPDSSQAQQLAAQAAKAAAGDDSAALADAQK	420
DB	617	DDAENETASILMSGFQMIHMENTENPDSSQAQQLAAQAAKAAAGDDSAALADAQK	676
QY	421	ALAEAALGKAGQOQOQILNALGQIASAAVVSAGVLPLOQVL	459



Db 677 ALEAALGKAGQQGILNALGQIASAAVVSAGVLPQQVL 715

## RESULT 2

D72042 conserved hypothetical protein CP0018 [imported] - Chlamydophila pneumoniae (strains CW1 N:Alternate names: chlpn 76 kda homolog\_1 (ct622); hypothetical protein CPn0728 C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000 C:Accession: D72042; D81623

R:Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A/2000; MUID:99206606

A:Accession: D72042

A:Molecule type: DNA

A:Residues: 1-651 <ARN>

A:Cross-references: GB:AF001654; GB:AF001363; NID:g4377031; PIDN:AAD18867.1; PID:g437703

A:Experimental source: strain CWL029

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255

A:Accession: D81623

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-651 <REA>

A:Cross-references: GB:AF002161; NID:g7188948; PIDN:AAF37914.1; PID:g718895

A:Experimental source: strain AR39, HL cells

C:Comment: This sequence was originally identified as homologous to part of a sequence (PIR:H71490).

C:Genetics:

A:Gene: Cpn0728; CP0018

Query Match 90.7%; Score 2188; DB 2; Length 651;

Best Local Similarity 98.9%; Pred. No. 3.5e-109;

Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNPPIGPGPIDERTPPADLSAAGLEASAAKSAEQAIRIAGAEAKPKESKTSVERWSI 60

DB 1 MNPPIGPGPIDERTPPADLSAAGLEASAAKSAEQAIRIAGAEAKPKESKTSVERWSI 60

QY 61 LRSVAVNLSLADKLGIASSNSSSTSRADVSTTATPTPPPTSDDYKTKQAOTAYDT 120

DB 61 LRSVAVNLSLADKLGIASSNSSSTSRADVSTTATPTPPPTSDDYKTKQAOTAYDT 120

QY 121 IFTSTSLADIQAALVSLQDAVNTIKDTAETAIAAEWETKNADAKVGAQITELAKY 180

DB 121 IFTSTSLADIQAALVSLQDAVNTIKDTAETAIAAEWETKNADAKVGAQITELAKY 180

QY 181 ASDNQAILDSLSGLKTSFDLLQALQSVANNKAAELKEMQDNVPVPGKTPAIQAQSLVD 240

DB 181 ASDNQAILDSLSGLKTSFDLLQALQSVANNKAAELKEMQDNVPVPGKTPAIQAQSLVD 240

QY 241 QTDTATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAKAAIATAKTOIAEAQ 300

DB 241 QTDTATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAKAAIATAKTOIAEAQ 300

QY 301 KKFPPDSPILQEAQEWVIAEQKDLKNIKPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMML 360

DB 301 KKFPPDSPILQEAQEWVIAEQKDLKNIKPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMML 360

QY 361 DDENETASILMSGFQRMHMFNTENPDSSQAQOELAAQAAKAAAGDSDSAAALADAQK 420

DB 361 DDENETASILMSGFQRMHMFNTENPDSSQAQOELAAQAAKAAAGDSDSAAALADAQK 420

QY 421 ALEAALGKAGQQGILNALGQIASAAVVSAGVLP 454

DB 421 ALEAALGKAGQQGILNALGQIASAAVVSAGVLP 454

## RESULT 3

E86581 CHLPN 76 kda homolog\_1 (CT622) [imported] - Chlamydophila pneumoniae (strain J138) C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001 C:Accession: E86581

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T. Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349

A:Accession: E86581

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-651 <STO>

A:Cross-references: GB:BA000008; NID:g8979100; PIDN:BA098935.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: CPJ0728

Query Match 90.7%; Score 2188; DB 2; Length 651;

Best Local Similarity 98.9%; Pred. No. 3.5e-109;

Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNPPIGPGPIDERTPPADLSAAGLEASAAKSAEQAIRIAGAEAKPKESKTSVERWSI 60

DB 1 MNPPIGPGPIDERTPPADLSAAGLEASAAKSAEQAIRIAGAEAKPKESKTSVERWSI 60

QY 61 LRSVAVNLSLADKLGIASSNSSSTSRADVSTTATPTPPPTSDDYKTKQAOTAYDT 120

DB 61 LRSVAVNLSLADKLGIASSNSSSTSRADVSTTATPTPPPTSDDYKTKQAOTAYDT 120

QY 121 IFTSTSLADIQAALVSLQDAVNTIKDTAETAIAAEWETKNADAKVGAQITELAKY 180

DB 121 IFTSTSLADIQAALVSLQDAVNTIKDTAETAIAAEWETKNADAKVGAQITELAKY 180

QY 181 ASDNQAILDSLSGLKTSFDLLQALQSVANNKAAELKEMQDNVPVPGKTPAIQAQSLVD 240

DB 181 ASDNQAILDSLSGLKTSFDLLQALQSVANNKAAELKEMQDNVPVPGKTPAIQAQSLVD 240

QY 241 QTDTATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAKAAIATAKTOIAEAQ 300

DB 241 QTDTATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAKAAIATAKTOIAEAQ 300

QY 301 KKFPPDSPILQEAQEWVIAEQKDLKNIKPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMML 360

DB 301 KKFPPDSPILQEAQEWVIAEQKDLKNIKPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMML 360

QY 361 DDENETASILMSGFQRMHMFNTENPDSSQAQOELAAQAAKAAAGDSDSAAALADAQK 420

DB 361 DDENETASILMSGFQRMHMFNTENPDSSQAQOELAAQAAKAAAGDSDSAAALADAQK 420

QY 421 ALEAALGKAGQQGILNALGQIASAAVVSAGVLP 454

DB 421 ALEAALGKAGQQGILNALGQIASAAVVSAGVLP 454

## RESULT 4

G71490

hypothetical protein CT622 - Chlamydia trachomatis (serotype D, strain UW3/Cx)

N:Alternate names: chlpn 76kda homolog CT622

C:Species: Chlamydia trachomatis

C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999

C:Accession: G71490

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchu

Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia

A:Reference number: A71570; MUID:99000809

A:Accession: G71490

A:Molecule type: DNA

A:Residues: 1-647 <ARN>

A:Cross-references: GB:AE001333; GB:AE001273; NID:g3329068; PIDN:AAC68226.1; PID:g33







Db 150 AADSARAASAGQAAS-SAGSASSAGTASTKATEASKSAAAAESSKSAATSAAGAKT 208  
 Qy 124 STSLADIAOAAVLSQDAVNTIKOTAADETAIAAEWETKNADAIVKGAQITELAKYASD 183  
 Db 209 SETNAS-----ASLSAASATSTATKASEAATSA-----RDAAAASKEAKSSETNASSSA 259  
 Qy 184 NOAILDSLGKLTSPDLLQTLQSVANNKAAELLEMQNDPVVPGTPTAQAOSLVQD 243  
 Db 260 SSAASATAAGNSAKAAKTS-----ETNARSSETAAGSASAAAGSKTAAASSASAAS 314  
 Qy 244 ATATQIEKDGNAIGDAYFAGQASGAVENAKSNNSISNIDSAAKAIATATQIAEAQKKE 303  
 Db 315 A-----GOASASATAAGKSA-----ESAASSASTATTKAGEATEQ- 349  
 Qy 304 PDSPILOAEQMVTOAEKDLNKPAGSDVPNPGTTVGGSKQOGSSIGSRVSMILLDDA 363  
 Db 350 -----ASAAARSASAAKTSETNAKASETSABESSKTAASASSASSASSASASK--DEA 402  
 Qy 364 ENETASILMSGFROMIHMFNTENPDQAQOELAAQA-----RAAKAAGDSDSAAALAD 417  
 Db 403 TROASAKSSATSTKATEAAGSATAAQAQSKSTAESAATRAETAAKRAEDIASVALED 462  
 Qy 418 AQKALEAALGKAGQOQILNALGQTASAAVYSAGVLPQQVLRIRYQAYVEOKLISEE 477  
 Db 463 ASTT-----KKGIV-----QLSSATNSTSETLAATP-----KAVKSAYDNAEKRLQK 504  
 Qy 478 DLNSA 482  
 Db 505 DONGA 509

RESULT 7  
 T31110  
 extracellular matrix binding protein - Abiotrophia defectiva (fragment)  
 C:Species: Abiotrophia defectiva  
 C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
 C:Accession: T31110  
 R:Manganelli, R.; van de Rij, I.  
 Infect. Immun. 67, 50-56, 1999  
 A:Title: Cloning and characterization of emb, a gene encoding the major adhesin of Streptococcus  
 A:Reference number: 20988; MUID:99081722  
 A:Accession: T31110  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2055 <MAN>  
 A:Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AAD03320.1  
 C:Genetics:  
 A:Gene: emb

Query Match 7.8%; Score 188; DB 2; Length 2055;  
 Best Local Similarity 23.9%; Pred. No. 0.034;  
 Matches 124; Conservative 68; Mismatches 197; Indels 130; Gaps 24;

Qy 19 ADLSAQGLEASAAKSAEQAQTAGAE-----KPKESKTSVERWSILRSVAVNLSIAD 73  
 Db 1602 AKKAKDAIDAATSNADVTQAQDAGNAINAVPQTPTAKTD-----AKNAVTAQAD 1651  
 Qy 74 KLGIASSNSSSTSRSDVSTTATPTPPPTSDDYKTAQATYDTFTSTSLADIAQ- 132  
 Db 1652 AKKDAIENDANITREKDAAKAKVDAEA-----TKAKNAID-----AATSNADVTAK 1698  
 Qy 133 -----ALVSLQDAVT-----NIKTAATDEETAI-----AAEWETKNADAIVKGAQITE 176  
 Db 1699 QNEGTAKINDVPQTPTAKTAKNAVDQAATDKKSAIENDPALTREKDAAKAVDAEATK 1758  
 Qy 177 LAKYASDNOAILDSLGKLTSPDLLQTLQSVANNKAAELLEMQNDPV--VPQKTPAI 234  
 Db 1759 -AKNAID-----AATSNADVTQAQDAGNAINAVP--QTPTA 1792  
 Qy 235 -----AQSLVDQ-----TDATATQIEKD--GNAIGDAYFAGQASGAVENAKSNNSI- 279  
 Db 1793 KTDKNAVDQAATDKKAAIENDPALTREKDAAKAVDA--EAKKAKDAIDAATSNADVT 1850

Qy 280 SNIDSAAKAI-----ATAKTQIAEA-----QKFPDPSPILOAEQMVIOAEKDLKN 325  
 Db 1851 AQDAGKDAINAVPQTPTAKTDAKNVDAQATDKKSAIENDPALTREKDAVRAKVDAEA 1910  
 Qy 326 IKPADGSDVPNPGTTVGGSKQOGS-SIGSI-RVSMILLDDAENETASTILMSGFROMIHMFN 383  
 Db 1911 KKAKDAIDAATSNADVTAKTEGTQAINAVPQTPTAKTDAKNVDAQATDKKNAIENDPA 1970  
 Qy 384 TENPDSQAQOELAAQARAARAKAGDSDSAAALADAQKALEAALGKAGQOQ--ILNALGQ 441  
 Db 1971 LTRREKDAAKAKKAKDAID--AATSNADV-----TAKQNEGTAKINDVPQ 2019  
 Qy 442 IASAAVYSAGVLPQQVLRIRYQAYVEOKLISEEDLN 480  
 Db 2020 TPTAKTDAKNVD-----QAATDKKSAIEND 2047

RESULT 8  
 S33441  
 EF protein - Streptococcus suis  
 C:Species: Streptococcus suis  
 C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 15-Oct-1999  
 C:Accession: S33441  
 R:Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.  
 submitted to the EMBL Data Library, May 1993  
 A:Description: Repeats in an extracellular protein of wek-pathogenic strains are abs  
 A:Reference number: S33441  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1822 <SMI>  
 A:Cross-references: EMBL:X71880; NID:g298031; PIDN:CAA50714.1; PID:g298032

Query Match 7.4%; Score 178; DB 2; Length 1822;  
 Best Local Similarity 22.9%; Pred. No. 0.098;  
 Matches 122; Conservative 80; Mismatches 201; Indels 130; Gaps 23;

Qy 13 TERPPADLSAQGLEASAAKSAEQAQTAGAEKPKES-KTDSVERWSIL-RSVAVALMS 70  
 Db 1201 TEKAKAAELAGEKSLTDTGKEARDAVELAKOLAKELAKTEEEETKIVKEKLAEDTRKA 1260  
 Qy 71 LADKLGIASSNSSSTSRSD-VDSSTTATPTPPPTSDDYKTAQATYDTFTSTSLAD 129  
 Db 1261 IEDPNLSDEKQAEIKKLTDAVAKTLATI-----RDNAKRTQEAERK-----QALAD 1309  
 Qy 130 IQAALVSLQDAVTNIKDTAATDEETAI--AEWETKNADAIVKGAQITELAKYASDQAI 187  
 Db 1310 LEKA-----KETOKIADKAAIDRLTILVKGELEATKODAKNKIAKDAKAAKEAIA 1364  
 Qy 188 LDSLGKLTSPDLLQTLQSVANNKA-----AELLKEMQNDPVVPGKTPAIAQSLVD 240  
 Db 1365 LTDEKKTFTD-----AVDAEAKANDAIASAATSPADVOKE-EDAGVA-----ATAEOVLD 1414  
 Qy 241 QTDATA-TQIEKDGNAIGDAYFAGQASGA-----VENAKSNNSISNIDSAAKAI 289  
 Db 1415 AAKQAKNKIAKDAKAAKEAIGSNPLNTDAEKFTTDAVDAEAKANDAI- 1465  
 Qy 290 ATAKTQIAEAQKPPDPSPILOAEQMVTOAEKDLNKPAGSDV-----PNFGTT--- 340  
 Db 1466 --AATSPADVOKE-EDAGVAAIADVDLDAKQDAK-NKIAKESDAKSAIDANPLNTDAE 1521  
 Qy 341 -----VGGSKQOGSSIGSRVSMI-----LDDAENETASTILM 372  
 Db 1522 KESAKKAVDAKAAKTAIDAITASTSPVEAQSAEDKGVGIAQDVLDAKQDAKNTAKEVA 1581  
 Qy 373 SGFROMIHMFNTENPDQAQOELAAQARAARAKAGDSDSAAALADAQKALEAALGKAGQ 432  
 Db 1582 AAKAIDAIDANPLNSDAEAKKAVDAKATTTDAID--ASTSPVEAQSAEDKGVGSI--R 1637  
 Qy 433 QGILNALGQIASAAVYSAGVLPQQVLRIRYQAYVEOKLISEED-LNSAVD 484







```

Query Match      7.0%; Score 170; DB 2; Length 1128;
Best Local Similarity 22.9%; Pred. No. 0.14;
Matches 117; Conservative 82; Mismatches 222; Indels 90; Gaps 20;

QY      10 IDEPRTPPADLS---AOGLEASAANKSAFAOARIAGAEA-KPKESKTDTSVERWSILRSA 64
      | | | | | : | : | | | | | | | | | | | | | | : | : |
Db      614 VAEAEKQAAEATKVAAEAEKQKAAEATKVAAEAEKQKAAEATKVAAEAEKQKAAEATKVAAE 673
      | | | | | : | : | | | | | | | | | | | | | | : | : |

QY      65 VNALMSLADLGLGIASSNSSSTSRSDVDSTTATPTPPPTSDDYTQAOATYDTFTS 124
      | | | | | : | : | | | | | | | | | | | | | | : | : |
Db      674 EKQAAEATKVAAEAEKQKAAEATKVAAEAEKQKAAEAT---KVAAEAEKQKAAEA---TK 725
      | | | | | : | : | | | | | | | | | | | | | | : | : |

QY      125 TSLADIQAAVLSDQDVTNIKDPAATDEBTAIAEWETKNADATKV-----CAQITEL 177
      | | | | | : | | | | | | | | | | | | | | | | : | : |
Db      726 VAEAEKQAAEATKVAAEAE-KQKAAE---EATKVAAEAEKQKAAEATKVAAEAEKQKAAEATKV 782
      | | | | | : | | | | | | | | | | | | | | | | : | : |

QY      178 AKYASDNCQAATLDSLGKLTISFDLLOTALLOSIV--ANNKNKAAELLK-----EMQDNPPWPGEKT 231
      | | | | | : | | | | | | | | | | | | | | | | : | : |

```

```

A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049
11:
Query Match 6.9%; Score 166.5; DB 2; Length 2232;
Best Local Similarity 19.7%; Pred. No. 0.52;
Matches 79; Conservative 70; Mismatches 174; Indels 79; Gaps
11:
QY 4 PIGPGPIDERTPPADLSAQ----GLEASAANKSAEAQRIAGAEAKPKESKTDVSVERWS 59
Db PSSOSPAPNTGSTTPSQTSOSPSPNPNPSSSTPTGSSQSTTPEGSTASSPTGSGTGF 657
QY 60 ILRSVAVNALMSL--ADKILGIASSNSSSSTSRSADVDSATTATAPTPPPPTSDDYKTQATA 117
Db 658 SVATEVTSQSTVSPGSSSLGTQSTNSPSPSPSLSPSTSGMSTLTSEPSS---TQSSGA 713
QY 118 YDTIFT--STSLLADTQAAALVSLQDAVTNKKPDAATDEETAAEAETWKNADAIKVCAQITE 176
Db 714 QSTLTTPSPNQSTLSLESTSGATTSGSGAGT-----MTSPGSSSVG----- 759
QY 177 LAKYASDNQATILDSGLKITSPLDLLQTLALQSVANNKAAELKEMQDNVPVPGKTPAIAQ 236
Db 760 -SSOGSTSPAAASTTSGEMTSQGSTQT-----PGSSSVSTSA 793
QY 237 SLVDDTATATQIEKDNGAICDAYFAGONASGA---VENAKSNNSISNIDSAKAALATAK 293
Db 794 ALLFTSQOSVSTNSPGSTVTRPSTVSGSTVTVTGSTEASTSGSVSSVSSPAPSTSQ 853
QY 294 TQIAEAQKFPDSPILQEAQOMVIAQEKDLKNKIPADGSDVPNPG-----TTVGGSKQQ 347
Db 854 N-----PNPSTSGSSSMITOSPYPSSQTSPTSPVSSSTTPSPGSPGTTTLTSTSPSPSO 903

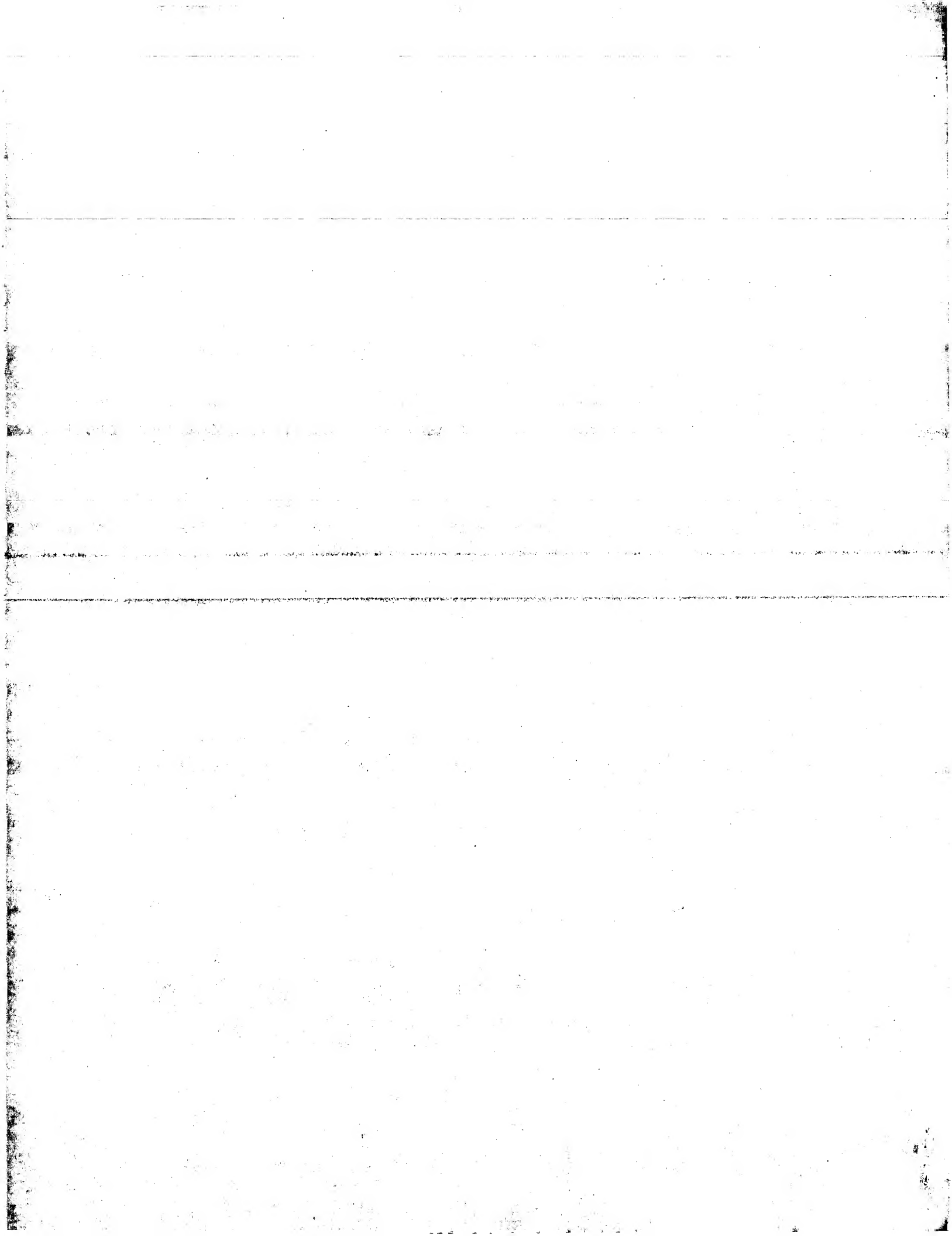
```

MOI: Biochem. Parasitol. 5/, 31/-330, 1993  
A;Title: Interruption of a Trypanosoma cruzi gene encoding a protein containing 14-amin



DECLASSIFICATION: 134022







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:42:39 ; Search time 76.51 Seconds  
(without alignments)  
234.816 Million cell updates/sec

Title: us-09-391-606-8  
Perfect score: 2412  
Sequence: 1 MVNPGPGIDERTPPAD.....QKLSEEDLNSAVDHHHHH 490

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190	7.9	1120	1 STFR_ECOLI	P76072 escherichia
2	174.5	7.2	545	1 HTR5_HALN1	Q48318 halobacteri
3	158	6.6	535	1 HTR4_HALN1	P33741 halobacteri
4	158	6.6	810	1 HTR4_HALN1	Q9hp84 halobacteri
5	158	6.6	641	1 HTR4_HALN1	Q48317 halobacteri
6	154.5	6.4	641	1 HRPK_PSESY	P41501 pseudomonas
7	154.5	6.4	763	1 HTR2_HALN1	Q9hp81 halobacteri
8	154.5	6.4	764	1 HTR2_HALN1	P71410 halobacteri
9	152.5	6.3	1528	1 SPAA_STRDO	P21979 streptococc
10	152	6.3	1561	1 SPAP_STRMU	P23504 streptococc
11	148	6.1	505	1 FLJB_SALTY	P52616 salmonella
12	148	6.1	881	1 YJH8_YEAST	P47033 saccharomyc
13	147	6.1	778	1 HTR6_HALSA	Q48319 halobacteri
14	146.5	6.1	774	1 STE_LAMBD	P03764 bacterioph
15	146	6.1	1714	1 SYEP_DROME	P28668 drosophila
16	145	6.0	705	1 CWBA_BACSU	Q02113 bacillus su
17	145	6.0	1565	1 PAC_STRMU	P11657 streptococc
18	144	6.0	1637	1 MRSP_STAAU	P80544 staphylococ
19	143.5	5.9	1306	1 MSB2_YEAST	P32334 saccharomyc
20	142.5	5.9	778	1 HTR6_HALN1	Q9hr92 halobacteri
21	140	5.8	436	1 Y868_CHLMU	Q9pj11 chlamydia m
22	139.5	5.8	2492	1 TALA_DICDI	P34633 dictyostell
23	139	5.8	797	1 VGLX_HSVEB	P28968 equine herp
24	138	5.7	1609	1 LMGI_HUMAN	P11047 homo sapien
25	138	5.7	2843	1 APC_HUMAN	P25054 homo sapien
26	137.5	5.7	2453	1 NCRI_MOUSE	Q60974 mus musculu
27	137	5.7	500	1 FLJB_SALAE	P52615 salmonella
28	137	5.7	535	1 HTR1_HALSA	P33955 halobacteri
29	137	5.7	1607	1 LMGI_MOUSE	P02468 mus musculu
30	136	5.6	1379	1 YFF9_SCHPO	O14066 schizosacch
31	136	5.6	2541	1 TALI_MOUSE	P26039 mus musculu
32	136	5.6	2845	1 APC_MOUSE	O61315 mus musculu
33	135.5	5.6	986	1 GM13_RAT	Q62839 rattus norv

34	135.5	5.6	1061	1 TRC4_ECOLI	P27189 escherichia
35	135.5	5.6	1969	1 MYSA_CAEEL	P12844 caenorhabdi
36	135.5	5.6	2090	1 HFC1_MESAU	P51611 mesocricetu
37	134	5.6	564	1 M12_STRPY	P19401 streptococc
38	134	5.6	1036	1 NIT2_NEUCR	P19212 neurospora
39	134	5.6	2842	1 APC_RAT	P70478 rattus norv
40	133	5.5	1972	1 MYHB_HUMAN	P35749 homo sapien
41	132.5	5.5	573	1 YEYB_YEAST	P40095 saccharomyc
42	132.5	5.5	1500	1 SSP5_STRGN	P16952 streptococc
43	132	5.5	2349	1 TPR_HUMAN	P12270 homo sapien
44	131.5	5.5	439	1 Y579_CHLTR	O84583 chlamydia t
45	131.5	5.5	1065	1 SED4_YEAST	P25365 saccharomyc

## ALIGNMENTS

RESULT 1  
STFR\_ECOLI STANDARD; PRT; 1120 AA.  
AC P76072: P77560;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE SIDE TAIL FIBER PROTEIN HOMOLOG FROM LAMBDOID PROPHAGE RAC.  
GN STFR OR B1372.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
{1}  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN {2}  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Aliba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
CC -!- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by: non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AE000234; AAC74454.1; ALT\_INIT.  
CC EMBL; D90774; BAA14966.1; -;  
CC EMBL; D90775; BAA14975.1; -;  
CC HSSP; P04002; IWFA.  
CC Ecogene; EG13370; stfr.  
CC InterPro; IPR000122; Chemotaxis\_transducer.  
CC Hypothetical protein; Fiber protein; Repeat; Complete proteome.  
SQ SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;



```

RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterheilt D.;
RA "A family of halobacterial transducer proteins.";
RL FEMS Microbiol. Lett. 139:161-168(1996).
CC
CC -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION.
CC
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005080; AAG19985.1; --
DR EMBL; X95589; CAA64842.1; -.
DR InterPro; IPR000122; Chemotaxis_transducer.
DR InterPro; IPRO03660; HAMP.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR Transducer; Transmembrane; Complete proteome.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
SQ SEQUENCE 545 AA; 57070 MW; FDD870389C2F428B CRC64;

Query Match 7.2%; Score 174.5; DB 1; Length 545;
Best Local Similarity 21.7%; Pred. No. 0.028;
Matches 100; Conservative 79; Mismatches 192; Indels 89; Gaps 17;

QY 12 ETERTPPADLSAGCLASANKSAEAQRIRAGAEAKPKESKTDSEVERSIILRSVAVNLMSL 71
DB 118 EEERA-----EAEERAEKAPQKAQEAI---QTAAESAQAKDAREARSAEIEQLAADLESQ 169
OV 72 ADKLGIASSNSSSTSRSDVDSTTATPTPPPT--SDDYKTKOATDYTI-----FTST 125

```

176	QY	SLADIQAALVSLQDVTNIKDT-----AATDEE-----TATAEWETKNADAIKVGAQITE	176
177	QY	SLADIQAALVSLQDVTNIKDT-----AATDEE-----TATAEWETKNADAIKVGAQITE	176
229	Db	ASREATAGAKEIODASQTVSESQVETAAAGTDDOREQLESVAEEMDSYSATVEEVAATAQS	288
177	QY	LKAYSADNOAILDSIGLKT-----SFDLLOTLQALQSVANNKAAELKEMODNPVYPGK	230
289	Db	VADTAADTTDVA-TAGQTAEDAIDAIDAQVETMOTTVANVDALEDLTLEIDD-----	340
231	QY	TPAIAQSLVDOTDAT-----ATQIEKDQNAIGD--AYFAGQNASGAVENAKSNNS	278
341	Db	---IABLISDIABOTNNLALNANIEAARAGSGGGSGNDGFVAVDEYKELATESORSAKD	397
279	QY	ISN-IPSAKAAIATAKTQIAEAOCKFPDPSILOEAEQOMVQIAERDLKNIKPADGSDVPNP	337
398	Db	IAELIEVQSOATATVEEI-----RVAEQRVNDGAAAVEETVDAFCGAVTENI	444
338	QY	GTTVGSQKQSGSIGSRVSNMLLDAAENETASILMSGFROMIHMFNTENPDSSQAQBELA	397
445	Db	QETTDGVQEE-----ISQAMDEQAQRSEVVSS-----VDDIATISQATA	483
398	QY	AQARAKAAGDDSAALADAQAALKAALGAGQOQOGLN	437
484	Db	DRAENVSAASEEQ-TASITTEVTSLSQSLAAQAADTLEDRLN	522

RESULT 3

HTRI_HALN1	STANDARD;	PRT;	535 AA.
ID	HTRI_HALN1		
AC	P33741; Q9HPF6;		
DT	01-FEB-1994 (Rel. 28, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		



20-AUG-2001 (Rel. 40, Last annotation update)  
SENSORY RHODOPSIN I TRANSDUCER (HTR-I) (METHYL-ACCEPTING PHOTOTAXIS  
PROTEIN I) (MPP-I).

GN HTR1 OR HTRI OR VNG1659G.  
OS Halobacterium sp. (strain NRC-1), and  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OC Halobacterium.  
OX NCBI\_TaxID=64091, 2242;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=NRC-1.  
RX MEDLINE=20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
RT "Genome sequence of Halobacterium species NRC-1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
RN [2]  
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-10; 350-377 AND 457-476.  
RP SPECIES=H.halobium; STRAIN=FLX5R;  
RC MEDLINE=93101637; PubMed=1465418;  
RX Yao V.J., Spudich J.L.;  
RT "Primary structure of an archaeobacterial transducer, a  
RT methyl-accepting protein associated with sensory rhodopsin I.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:11915-11919(1992).  
CC -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR  
CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.  
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE005075; AAG19913.1; -;  
DR EMBL; L05603; AAA72315.1; -;  
DR PIR; A47190; A47190.  
DR InterPro; IPR000122; Chemotaxis\_transducer.  
DR InterPro; IPR000658; DUF5.  
DR Pfam; PF00672; DUF5; 1.  
DR Pfam; PF00015; MCPsignal; 1.  
DR SMART; SM00304; HAMP; 2.  
DR SMART; SM00283; MA; 1.  
KW Transducer; Photoreceptor; Transmembrane; Methylation.  
FT INIT\_MET 0  
FT DOMAIN 1 13  
FT TRANSMEM 14 28  
FT DOMAIN 29 38  
FT TRANSMEM 39 54  
FT DOMAIN 55 535  
FT MOD\_RES 265 265  
FT MOD\_RES 272 272  
FT MOD\_RES 279 279  
FT MOD\_RES 463 463  
FT MOD\_RES 472 472  
FT SEQUENCE 535 AA; 56544 MW; B5945E4F66A9D091 CRC64;  
Query Match 6.69; Score 158; DB 1; Length 535;  
Best Local Similarity 18.79; Pred. No. 0.19;  
Matches 88; Conservative 83; Mismatches 189; Indels 110; Gaps 15;

Qy 8 GPIDETERTPPADLSAQGLEASAAKSAE--AQLIAGAEAKPKE---SKTDSVERW--- 58  
D 103 GRNEMERT-RADLEETQAEAREAEAEQAEQAEAREAREARELAAYQDTAKRYGETM 161  
Qy 59 -----SILRSVNALMSLADKLGTIASSNSSST 86  
D 162 EAAATGDLTORVDVDTHEAMETVGTAFNMMDLQATVTVTTVADEIEAKTERMSSET- 220  
Qy 87 SRSADVST---TATAPTPPTSDYKTAQTAQVDTFTSTSLADIQAALVSLQDAVTN 143  
D 221 --SADIEASAGDTVEANSKIESQANDQTEILDSAD-----DVQVSAEAEIAAT 269  
Qy 144 IKDAAATDEETAIAE-----WETKNADAIKVGQITELAKIYASDNOAIL 188  
D 270 IDDLASRESVATASDAARDSSKALDEMSSTIETEDVDAQVQELRQDVAEITDIVDI 329  
Qy 189 DSLGKLTSTFDLLOLALLOSVAANN-----KAELLKEMQDNVPVPGKTPAIQAS 237  
D 330 TDIGETQTNMLALNASIEAARAGGNADGGSFVVADEVKDLAETQDR---ANEIAAVVEK 386  
Qy 238 LVQDQTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAATATAKTOIA 297  
D 387 VTAQTEDVTASIQOTRIV-----ESGSETVE-----STLRDITADSLAEVNSID 434  
Qy 298 EAQKFPDSPILQEAQEMVIOAEKDLKNIPADGSDVPNPGTTVGSKQOQSSIGSIRVS 357  
D 435 EIQRSTSEQAEVQSTATSVERVAGLSDDTTALASDAE---SAVIGQRESAEIEAA--- 487  
Qy 358 MLLDDAENETASILMSGFQMHMTENPDSQAQOQLAAQARAAGAAG 407  
D 488 -SLEQFQNTAEVQEQS-----RVASFTVATEDSETAGGSVEQPMRAGADG 532  
RESULT 4  
HTR4\_HALN1 STANDARD; PRT; 810 AA.  
ID Q9HP84;  
AC 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DE HALOBACTERIAL TRANSDUCER PROTEIN IV.  
GN HTR5 OR HTPIV OR VNG1760G.  
OS Halobacterium sp. (strain NRC-1).  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OX NCBI\_TaxID=64091;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
RT "Genome sequence of Halobacterium species NRC-1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
CC -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL  
CC TRANSDUCTION (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE005080; AAG19986.1; -;  
DR InterPro; IPR000122; Chemotaxis\_transducer.



```
DR InterPro: IPR000658; DUF5.
DR InterPro: IPR003660; HAMP.
DR Pfam: PF00672; DUF5; 1.
DR SMART; SM00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
KW Transducer; Transmembrane; Complete proteome.
FT DOMAIN 1 38 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 39 59 POTENTIAL.
FT DOMAIN 60 323 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 324 344 POTENTIAL.
FT DOMAIN 345 810 CYTOPLASMIC (POTENTIAL).
FT SEQUENCE 810 AA; 85219 MW; 4BF36E4B7D22BD80 CRC64;

Query Match 6.68; Score 158; DB 1; Length 810;
Best Local Similarity 22.08; Pred. No. 0.31;
Matches 102; Conservative 74; Mismatches 169; Indels 118; Gaps 22;

QY 27 EASAANKSA-BAORTAGAEAKPKESKTDSEVERWSILRSVAVNALMSLADKLGIASSNSSSS 85
Db 398 EAAREAQATEAQDAEAE---RERAEADAREADAKADAPAL-----AAELEAQ 444
QY 86 TSRSADVSTTATPTPPPTSDDYKTAQATYDTFTSTSLADIQAALVSLQDAVNTNIK 145
Db 445 AERYSDVMAACADGDLTRMPADDDTNEAMAAIAASF-NEMLAQWEHTIIDIQE---FA 499
QY 146 DTAATDEETATAEWEKTKNADAIKVAQIT----ELAKYASDNQAALDSL-GKLTSPDLL 200
Db 500 DAVATASEEA-----EVGAADAERASQGVSEVQIEAGADEQRNMLDVTSGEMT--DL- 551
QY 201 QTALQSVANNKAAELKEMQDNPVPGKTPAJAQSLVDOTDATATQIEKDGNAGDAY 260
Db 552 -SAAIEVA-----ASDSVAESHQTA-EIARDGE-----580
QY 261 FAGQNASGAVENAKNSNISIDSAKAIAIATAKTOIAEPAQKKFPD--SPILQEAQMVQ 318
Db 581 ---QTAEDAIE--RSLTVQEAIDATVQNVREALDDQMAEI-SEIVDLISDIAEQTMLALN 634
QY 319 AE-----KDLKNKPKADGSDVPNPVPGKTPAJAQSLVDOTDATATQIEKDGNAGDAY 358
Db 635 ANIEAARADKSGDGFVAVVADVEKDLAEETOESAGDIERRITEV--QSOTTAIVAEARAE 692
QY 359 LDDAENETASILMSGFRQMI-HMFNT-----ENPDQAAQOE-----395
Db 693 ESMADGIDAVEEVVDATFVSDHADETTDTGVQIEISDITDDQAASTEEAVSMTEEVADLS 752
QY 396 -LAAQARAAGDSDSAAALADAKALEALGRAGQOQGIIN 437
Db 753 STAGEAQSVSAAREEQ-AASMSEISDSVESLSGQAEQLKALLS 794

RESULT 5
HTR4_HALSA
ID HTR4_HALSA STANDARD; PRT; 810 AA.
AC Q4B317;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HALOBACTERIAL TRANSDUCER PROTEIN IV.
GN HTR5 OR HTPIV.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S9;
RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterhelmt D.;
RT "A family of halobacterial transducer proteins.";
RL FEMS Microbiol. Lett. 139:161-168(1996).
```











RT RT "Regions of the Streptococcus sobrinus spaA gene encoding major  
RL J. Bacteriol. 172:3988-4001(1990).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
CC -1- MISCELLANEOUS: IMMUNODOMINANT DETERMINANTS ARE LOCATED IN THE  
CC C-TERMINAL TWO-THIRDS OF THE SPAA PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE SPAP/SSPS/SPAA FAMILY.  
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS  
CC IN THE REGION OF THE MEMBRANE ANCHOR.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X57841; CAA40973.1; -  
DR EMBL; M38210; AAA26977.1; -  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; 1.  
KW Antigen; Signal; Transmembrane; Repeat; Cell wall.  
FT SIGNAL 1 ? POTENTIAL.  
FT CHAIN ? ? CELL SURFACE ANTIGEN I.  
FT CHAIN ? ? CELL SURFACE ANTIGEN II.  
FT DOMAIN 1503 1508 CONSERVED IN GRAM-POSITIVE COCCI SURFACE  
FT PROTEINS.  
FT CONFLICT 427 427 A -> E (IN REF. 2).  
FT CONFLICT 431 431 Q -> K (IN REF. 2).  
FT CONFLICT 434 434 A -> K (IN REF. 2).  
FT CONFLICT 531 531 A -> S (IN REF. 2).  
FT CONFLICT 600 600 A -> S (IN REF. 2).  
SQ SEQUENCE 1528 AA; 165572 MW; B235F9CCD92838E0 CRC64;

Query Match 6.3%; Score 152.5; DB 1; Length 1528;  
Best Local Similarity 23.3%; Pred. No. 1.3;  
Matches 116; Conservative 72; Mismatches 207; Indels 103; Gaps 24;

QY 38 ORTGAERAKPKESKTSVERWSILRSVNALMSLADKLGIASSNSSSTS--RSADVST 95  
DB 16 KLLCA-AKVSGRLS---GALLGTAI--LASGAGQALAEETSTTSVGGDTAVVGT 69  
QY 96 TATATPPPPPTSDDYKTAQTAQY-----DTFTSTSLADIQAALVSLQDAVTNKDTA 148  
DB 70 TGNPATNLDPKODNPSQAETSOQAQARKTGAMSDVSTSELDAAKSPQEAQVTVSQA 129  
QY 149 ATDEETATAAEWETNDAIK--VQAQITELAKYASD-----NQALDSLGLKLTSD 199  
DB 130 TVNKGTVPEPDEANQKEPEIKDDYSQAADIQKATEDYKASVAANOQETDRINQEA 189  
QY 200 LOTALQSVANNHKAEE--LKEMODNPVPGKTPAIAQSLVDQTDATATQTEKDGNAIG 257  
DB 190 AQYE--QDLAANKAEVERSLMRPRPIYEAK---LAQNKDL--AAIQANSQSQA-- 240  
QY 258 DAYFAGQN-----ASGAVENAKSNISNIDSAKAAIATAKTQIAEAKK 302  
DB 241 -AYAAAKEAYDKWARVOAANAANKAYEEAALAAANTAKN-DOIKAEIEAIQORSKA 295  
QY 303 FPDSPILQAEQVMIOAEKDLNKPAGSDVDPNPGTVGGSKQOOGSSIGSRVSMLLDD 362  
DB 296 -----DYEAKLAQYKDLAAAGAGNANEDYQAKAAVEQELARY-----Q 337  
QY 363 AENETASTILMSGFQMIHNTFN-----PDSQAQQLAAQAAKAAAGDSDSAAALADA 418  
DB 338 AANAANK---QAYEQALAAANSKNAQITAENEAIQON--AQAKA-----DYEAKLAQY 385  
QY 419 QXALEEA--LGKAGQOQGLNALG-----QIASAAVVSAGVLPLOQVLWIRARYQAY 468  
DB 386 QRDLAQAQSGNAANEADYQEKLAAYEKELARYQAANAANKAYE---QQVQAANAANKAEI 442  
QY 469 VE-QKLSEEDLNSAVDH 485

DB 443 TEANRAIRERNAKAKTDY 460  
RESULT 10  
SPAP\_STRMU STANDARD; PRT; 1561 AA.  
ID P23504;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE CELL SURFACE ANTIGEN I/II PRECURSOR.  
GN SPAP.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_taxid=1309;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-NG5 SEROTYPE C;  
RX MEDLINE=90076473; PubMed=2687020;  
RA Kelly C., Evans P., Bergmeier L., Lee S.F., Proquleske-Fox A.,  
RA Harris A.C., Aitken A., Bleiweis A.S., Lehner T.;  
RT "Sequence analysis of the cloned streptococcal cell surface antigen  
RT I/II.";  
RL FEBS Lett. 258:127-132(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NG5 SEROTYPE C;  
RX MEDLINE=91207143; PubMed=1982405;  
RA Kelly C., Evans P., Ma J.K.C., Bergmeier L.A., Taylor W., Brady L.J.,  
RA Lee S.F., Bleiweis A.S., Lehner T.;  
RT "Sequencing and characterization of the 185 kDa cell surface antigen  
RT of Streptococcus mutans.";  
RL Arch. Oral Biol. 35:338-385(1990).  
CC -1- FUNCTION: SURFACE PROTEIN ANTIGEN IMPLICATED IN DENTAL CARIES.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS  
CC IN THE REGION OF THE MEMBRANE ANCHOR.  
CC -1- SIMILARITY: BELONGS TO THE SPAP/SSPS/SPAA FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X17390; CAA35253.1; -  
DR PIR; S06839; S06839.  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; 1.  
KW Antigen; Signal; Transmembrane; Repeat; Cell wall; Dental caries.  
FT SIGNAL 1 38  
FT CHAIN 39 ? ? CELL SURFACE ANTIGEN I.  
FT CHAIN 997 1561 CELL SURFACE ANTIGEN II (PROBABLE).  
FT DOMAIN 39 1536 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1537 1556 MEMBRANE ANCHOR.  
FT DOMAIN 1557 1561 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 60 550 HELICAL (POTENTIAL).  
FT DOMAIN 219 464 3 X TANDEM REPEATS, ALA-RICH.  
FT DOMAIN 847 963 3 X TANDEM REPEATS, PRO-RICH.  
FT DOMAIN 1528 1533 CONSERVED IN GRAM-POSITIVE COCCI SURFACE  
FT PROTEINS.  
SQ SEQUENCE 1561 AA; 170060 MW; 540D92768FC8B4B CRC64;

Query Match 6.3%; Score 152; DB 1; Length 1561;  
Best Local Similarity 22.3%; Pred. No. 1.4;  
Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;



QY 48 KESKTSVERWSILRSVAVNALMSLADKLGIASSN-SSSSTSRSDVSTTATATPTPPPT 106  
 DB 2 KVKTYGFRKSKISKTLCGVLGVAASVAGQVFADEITTTSDVTKVGTGTGNPAT 61  
 QY 107 -----SDDYKQATAYTITST-----SLADIAALVS-LQDVTNKT 147  
 DB 62 NLPEAGSQAQESQTKLERQMVHTTEVPKTDLDQAADKAGSVNVQDADVN-KGT 120  
 QY 148 AATDEETAAEWE-----TKNADAIK-----VCAQITELAKYASDQALDSIGK 193  
 DB 121 VKTAE-AYOKETEIKEDYTOAEDIKTTDOYKSDVAHAEAEVAKIKAKNQAEQYQK 179  
 QY 194 LTSFOLL-QTALLOSVAANNKAEL-----LKEMQDNPVPGKTPAIAQSLVDQTDAT 247  
 DB 180 ----DWVAKHAEVERINAANAASKTAYEAKLAQYQADLAQVOKTNAANQASY-QKALAA 234  
 QY 248 QTE-----KQNAIGDAYFAGNAGAVENAKSNNSISNIDSAKAIATAK-----TQTAEA 299  
 DB 235 QAEKRVQEAANAAKAAAY---DTAVANNAKNTETAANAEBIRKRNATAKAEYETKLAQY 291  
 QY 300 Q---KKFPDSPLOEAE-OMVIOA-EKDLKNIKPAD-----GSDVPNPGTTVG 342  
 DB 292 QAEKRVQEAANAANEAQYQAKLTAYOTELARVORANADAKAAYEAANAANKNAALTAE 351  
 QY 343 GS-----KQGGSIGSIRVSMLLDDAENETASILMSGFROHMFNTE 385  
 DB 352 NTAIKORNENAKATYEAALKOYEADLAQVAKANAANEADYQAK---LTAYOTELARVQKA 408  
 QY 386 NPDSQAQOELAAQARAQAA--GDSAAAL-ADAQKALEAALGK 428  
 DB 409 NADAKAAYEAANAANAALTAENTAIAKRNADAKADYEAALAK 454

# RESULT 11 FLJB\_SALTY

ID FLJB\_SALTY STANDARD; PRT; 505 AA.  
 AC P52616; P97159;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PHASE-2 FLAGELLIN.  
 GN FLJB OR H2.  
 OS Salmonella typhimurium.  
 OC Bacteria; proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SL 375;  
 RX MEDLINE=95325331; PubMed=7541401;  
 RA Vanegas R.A., Joys T.M.;  
 RT "Molecular analyses of the phase-2 antigen complex 1,2... of  
 RT Salmonella spp.";  
 RL J. Bacteriol. 177:3863-3864(1995).  
 RN [2]  
 RP SEQUENCE OF 1-37 FROM N.A.  
 RX MEDLINE=82049491; PubMed=6271461;  
 RA Silverman M., Zieg J., Mandel G., Simon M.;  
 RT "Analysis of the functional components of the phase variation  
 RT system.";  
 RL Cold Spring Harb. Symp. Quant. Biol. 45:17-26(1981).  
 RN [3]  
 RP SEQUENCE OF 482-505 FROM N.A.  
 RC STRAIN-SJ2353;  
 RA Minorance J., Tanaka S., Tominaga A., Enomoto M.;  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO  
 CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.  
 CC -!- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE  
 CC BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLIN, TERMED  
 CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics, and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC  
 DR EMBL: U17177; AAC43354.1; -  
 DR EMBL: V01370; CAA24655.1; -  
 DR EMBL: D26168; BAA05156.1; -  
 DR StyGene; SG10564; fljB.  
 DR InterPro; IPR001029; Flagellin\_C.  
 DR InterPro; IPR001492; Flagellin\_N.  
 DR Pfam; PF00700; Flagellin\_C; 1.  
 DR Pfam; PF00669; Flagellin\_N; 1.  
 DR PRINTS; PR00207; FLAGELLIN.  
 DR ProDom; PD000316; Flagellin\_C; 1.  
 DR Flagella.  
 KW Flagella.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT CONFLICT 37 37 I -> S (IN REF. 2).  
 SQ SEQUENCE 505 AA; 52404 MW; 3A5CC404AF7AF88B CRC64;

Query Match: 6.18; Score 148; DB 1; Length 505;  
 Best Local Similarity 21.0%; Pred. No. 0.56;  
 Matches 93; Conservative 59; Mismatches 168; Indels 122; Gaps 20;

QY 23 AOGLEASNAKSAEAKRIAG-----AEAPKESKTSVERWSILRSVAVNALMSLADKL 75  
 DB 32 SSGLRINSAKDAAGQAIAANFTANIKGLTQASRNANDGISIAQTTEGALNEINNLRV 91  
 QY 76 GIASSSSSTSRSDVSTTATATPTPPPTSDDYKTAQ-----TAYDTIFT----- 123  
 DB 92 RELAVQSANSTNSQSDLSIOAEI-TQRLNEIDRVSGTQFNGVKVLAQDNTLTIQVGAN 150  
 QY 124 -----STSLADIQAALVSLQ-----DAVTNINKDTAATD-----EETAAAE 159  
 DB 151 DGETIDIDLKQINSQTILGSLNVQKAYDKVDTAVTTKAYANNGTTLDSGLDDAAIKAA 210  
 QY 160 WETKNADAIVGAOITELAKYASDQALDSIGKLTSPDLLQ-----TALLOS 207  
 DB 211 TGGTNGTASVTGGAV-----KFDADNNKYFVTIGGTGADAANKGDYEVNVAQDGTVLAA 266  
 QY 208 VANN-----KAAELKEMODNPV-----PGKTPAIAQSLVDQTDATATQI-----EKD 252  
 DB 267 GATKTPAGATTKTEVQELKDTPAVVSADAKNALIAGG-VDATDANGAELVKMSYTDKN 325  
 QY 253 GNAL-----GDAYFAG--ONASGAVENAKSN-----NSISNIDSAKAAI 289  
 DB 326 KKTIEGGYALKRAGDKYAAADYDEATGAIK-AKTTSTYTAADGTTKTAANQLGGVD-GKTEV 383  
 QY 290 AT-----AKTQIAEAKQKFPDPSILQEAQEMVIOAE-----KDLKNIKPA 329  
 DB 384 VTIDGKTYNASKAGHDFKAQPELAEEAAKTENP-LOKIDAAALQVDAALSDILGAVQNR 442  
 QY 330 GDSVPNPGTTVGSKQGGSSI 351  
 DB 443 FNSAITNLGNTVNNLSEARSRI 464

## RESULT 12

YJH8\_YEAST  
 ID YJH8\_YEAST STANDARD; PRT; 881 AA.  
 AC P47033;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL 89.2 KDA PROTEIN IN SCPI60-SMC3 INTERGENIC REGION.  
 GN YJL078C OR J1027.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.







QY 238 LVDTDTATATQIEKDN--AIGDAY--FAGQNASGAVE-----NAKSNNSISNIDS 284  
 Db 603 LALNANTEAARADODGDFAVADVKDLADESKASAAEALVAEVRAGTQETSVAAADR 662  
 QY 285 AKAAIATAKTQIAEAAQKFFDPSILOEAEQWVIOAE---KDLKNIRKPADGSDVNPFGTTV 341  
 Db 663 IOERVSGVETVSTERS-----LSEIAGRIAEADTGVGEISNAMDDQAAVSVDYTTAV 716  
 QY 342 GGSQKQSSIGSIRVSMMLDDAENETASILMSGFROMIHMFENTEN-PDSQAAQOEEAAQA 400  
 Db 717 GDVAALGEETATEASTA--DAAAEQASTLSDVAAQ-----TETLAHAVALREHAAQF 768  
 QY 401 RAA 403  
 Db 769 EVA 771

RESULT 14  
 ID STP\_LAMB STANDARD; PRT; 774 AA.  
 AC P03764; P03745;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SIDE TAIL FIBER PROTEIN.  
 GN STP.  
 OS Bacteriophage lambda.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;  
 OC Lambda phage group.  
 OX NCBI\_TaxID=10710;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83189071; PubMed=62211115;  
 RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;  
 RT "Nucleotide sequence of bacteriophage lambda DNA.";  
 RL J. Mol. Biol. 162:729-773(1982).  
 RN [2]  
 RP IDENTIFICATION AS STP.  
 RX MEDLINE=92165720; PubMed=1531648;  
 RA Haggaard-Ljungquist E., Hailing C., Calendar R.;  
 RT "DNA sequences of the tail fiber genes of bacteriophage P2: evidence  
 RT for horizontal transfer of tail fiber genes among unrelated  
 RT bacteriophages.";  
 RL J. Bacteriol. 174:1462-1477(1992).  
 RN [3]  
 RP RECONSTRUCTION OF STP.  
 RX MEDLINE=93068310; PubMed=1439823;  
 RA Hendrix R.W., Duda R.L.;  
 RT "Bacteriophage lambda P2: not the mother of all lambda phages.";  
 RL Science 258:1145-1148(1992).  
 CC -1- MISCELLANEOUS: THE COMMON LABORATORY STRAIN OF BACTERIOPHAGE  
 CC LAMBDA; LAMBDA P2; CARRIES A FRAMESHIFT MUTATION RELATIVE TO UR-  
 CC LAMBDA, THE ORIGINAL ISOLATE. THE UR-LAMBDA VIRIONS HAVE THIN,  
 CC JOINTED TAIL FIBERS (SIDE TAIL FIBERS) THAT ARE ABSENT FROM LAMBDA  
 CC WILD TYPE. RELATIVE TO LAMBDA P2, UR-LAMBDA HAS EXPANDED  
 CC RECEPTOR SPECIFICITY AND ADSORBS TO E. COLI CELLS MORE RAPIDLY.  
 CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.  
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS  
 CC CORRECTED IN POSITION TO 396 TO RECREATE THE ORIGINAL STP PROTEIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: J02459; AAA96555.1; ALT\_FRAME.  
 CC EMBL: J02459; AAA96557.1; ALT\_FRAME.  
 CC PIR: A04389; QXBPL1.  
 CC PIR: A04370; QXBP2L.

DR InterPro; IPR000122; Chemotaxis\_transducer.  
 DR InterPro; IPR001778; POA\_allergen.  
 KW Fiber protein.  
 SQ SEQUENCE 774 AA: 77527 MW: CDD1DF85B919123B CRC64;  
 Query Match 6.1%; Score 146.5; DB 1; Length 774;  
 Best Local Similarity 20.9%; Pred. No. 1.1;  
 Matches 90; Conservative 66; Mismatches 209; Indels 65; Gaps 13;  
 QY 26 LEASAANKSAEORIAAGAEAKPKESKTSVERVSIILRSVAVNMLSLADKLGITASSNSSS 85  
 Db 110 VEEVARNASVQSTADAKKSGAGSASAAQVAALVTDATDSARAASSTAGQAAS-SAGE 168  
 QY 86 TERSADVSTTATPTPPPTSDDYKTAQTAQVDTFTTSTSLADIQAALVSLQDAVNIK 145  
 Db 169 ASSGAEEASAKATEAEKSAEASAAEESKNAATSAAGAAKTSET-----NAAASQGSAAATSAS 223  
 QY 146 DTAATDEETAIAAEWETKNADAIVGAQITELAKYASDQAILDLSLGLKLTSLFDLLQTLALL 205  
 Db 224 TAATKASEAATSA---ROAVASKEAAKSSSETNASSAGRAASSATAAENSARAAKTS-- 277  
 QY 206 QSVANNKAAELKEMQDNVPVPGKTPAIAQSLVDOTDATAIOIEKDGNAIGDAYFAGQN 265  
 Db 278 ---ETNARSSETAERSASAAADAKTAAGSASTATKAT-----EAAGSAVSASOS 326  
 QY 266 ASGAVENA-KSNNSISNIDSAAKAAIA-----TAKTOIAEAAQKKFPDPSILQEAQWVIOA 319  
 Db 327 KSAEAAAIRAKKSAKRAEDIASAVALEADATTKRGIVQLSSATNSITSETLATPKAVK 386  
 QY 320 EKDLKNIK-PADG---SDVPNPGTIVGGSKQKQSSIGSIRVSMMLDDAENETASILMSGF 375  
 Db 387 VMDETNRKAPLDSPALTGTPAPTALRGTT-----NNTQIANTAFVLAAL 430  
 QY 376 ROMIHMFENTENPDSQAAQOELAAQAAKAAAGDDSAALADAAKALEAALGALGAGQOOGI 435  
 Db 431 ADVI-----DASPDALNTLNELAA-----ALGND-----PDPATTMTNALAGKQPRNAT 474  
 QY 436 LNALGOIASA 445  
 Db 475 LTALAGLSTA 484

RESULT 15  
 ID SYEP\_DROME STANDARD; PRT; 1714 AA.  
 AC P28668; Q9VCF5;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DE BIFUNCTIONAL AMINOACYL-TRNA SYNTHETASE [INCLUDES: GLUTAMYL-TRNA  
 DE SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE); PROLYL-TRNA  
 DE SYNTHETASE (EC 6.1.1.15) (PROLINE--TRNA LIGASE)].  
 GN AATS-GLUPRO OR CG5394.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92097547; PubMed=1756734;  
 RA Cerini C., Kerjan P., Astier M., Gratecos D., Mirande M., Semeriva M.;  
 RT "A component of the multisynthetase complex is a multifunctional  
 RT aminoacyl-tRNA synthetase.";  
 RL EMBO J. 10:4267-4277(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=OREGON-R;  
 RX MEDLINE=97217441; PubMed=9063462;  
 RA Cerini C., Semeriva M., Gratecos D.;  
 RT "Evolution of the aminoacyl-tRNA synthetase family and the  
 RT organization of the Drosophila glutamyl-prolyl-tRNA synthetase gene.



RT Intron/exon structure of the gene, control of expression of the two  
 RL mRNAs, selective advantage of the multienzyme complex.;  
 RP Eur. J. Biochem. 244:176-185(1997).  
 RN [3].

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RD MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.A.,  
 RA Brandon R.C., Rogers J., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Held G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*;"  
 RL Science 287:2185-2195(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +  
 CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(LEU).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-PROLINE + TRNA(PRO) = AMP +  
 CC PYROPHOSPHATE + L-PROLYL-TRNA(LEU).  
 CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS  
 CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE  
 CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,  
 CC ARGINYL, AND ASPARTYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY  
 CC PROTEINS, P18, P48 AND P43.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I  
 CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II  
 CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 6 "WHEP-TRS" DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M74104; AAC28594.1; -  
 CC DR EMBL; U59923; AAC47469.1; -  
 CC DR EMBL; AE003745; AAF56211.1; -  
 CC DR PIR; S18644; S18644.  
 CC HSP; P00962; IQRU.

DR FlyBase: FBgn0005674; Aats-glupro.  
 DR InterPro: IPR002106; AA-TRNA\_ligase\_II.  
 DR DR InterPro: IPR000738; WHEP-TRS.  
 DR DR InterPro: IPR000924; TRNA-synt\_1c.  
 DR DR InterPro: IPR002314; TRNA-synt\_2b.  
 DR DR InterPro: IPR001412; TRNA-synt\_I.  
 DR DR InterPro: IPR002316; TRNA-synt\_pro.  
 DR DR Pfam: PF00749; TRNA-synt\_1c; 1.  
 DR DR Pfam: PF00587; TRNA-synt\_2b; 1.  
 DR DR Pfam: PF00458; WHEP-TRS; 6.  
 DR DR PRINTS: PR00987; TRNASYNTHGLU.  
 DR DR PRINTS: PR01046; TRNASYNTHPRO.  
 DR DR PROSITE: PS00178; AA-TRNA\_LIGASE\_I; 1.  
 DR DR PROSITE: PS00179; AA-TRNA\_LIGASE\_II\_1; FALSE\_NEG.  
 DR DR PROSITE: PS00339; AA-TRNA\_LIGASE\_II\_2; FALSE\_NEG.  
 DR DR PROSITE: PS00762; WHEP-TRS; 6.  
 DR DR Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 DR KW Multifunctional enzyme; Repeat.  
 DR FT DOMAIN 170 754 GLUTAMYL-TRNA SYNTHETASE.  
 FT DOMAIN 755 800 WHEP-TRS 1.  
 FT DOMAIN 827 872 WHEP-TRS 2.  
 FT DOMAIN 901 945 WHEP-TRS 3.  
 FT DOMAIN 980 1025 WHEP-TRS 4.  
 FT DOMAIN 1055 1100 WHEP-TRS 5.  
 FT DOMAIN 1129 1173 WHEP-TRS 6.  
 FT DOMAIN 1174 1180 POLY-GLY.  
 FT DOMAIN 1207 1714 PROLYL-TRNA SYNTHETASE.  
 FT SITE 209 220 "HIGH" REGION.  
 FT SITE 438 442 "KMSKS" REGION.  
 FT BINDING 441 441 ATP (BY SIMILARITY).  
 FT CONFLICT 102 106 TSPLP -> DKST (IN REF. 3).  
 FT CONFLICT 233 234 VC -> AF (IN REF. 3).  
 FT CONFLICT 341 345 NTACA -> KYCVR (IN REF. 3).  
 FT CONFLICT 583 583 K -> R (IN REF. 3).  
 FT CONFLICT 692 692 L -> A (IN REF. 3).  
 FT CONFLICT 753 753 T -> S (IN REF. 3).  
 FT CONFLICT 802 802 T -> S (IN REF. 3).  
 FT CONFLICT 873 873 P -> T (IN REF. 3).  
 FT CONFLICT 887 887 G -> V (IN REF. 3).  
 FT CONFLICT 1201 1201 P -> PA (IN REF. 3).  
 FT CONFLICT 1461 1461 MISSING (IN REF. 3).  
 FT CONFLICT 1587 1587 G -> V (IN REF. 3).  
 SQ SEQUENCE 1714 AA; 189197 MW; 6F8C58045E48A8C CRC64;

Query Match 6.1%; Score 146; DB 1; Length 1714;  
 Best Local Similarity 19.8%; Pred.No. 3.2;  
 Matches 101; Conservative 74; Mismatches 192; Indels 142; Gaps 20;

QY 7 PGFI-----DETERTPPADLSAQGLEASAAKSAEAQRIAG-----A 43  
 DB 701 PSFIVLFSIPDGHGTHKDVPTSGLVNAPDAKATKASSPVSSGQASELDSQITQGGDLVR 760  
 QY 44 EAPKESKTSVSWKSTLRSAVNALMSADKLGLTASSNSSSTSRSDVSTTAT-APTTP 102  
 DB 761 DLAKSKAAKQDID-----VAVKKLLAL-----KADYKSATGKDWKPGQTSAPVP 807  
 QY 103 PPPTSDDYKTAQATYDTFTSTSLADIAQALVSLQDAVTNPKDTAATDEATA----- 157  
 DB 808 AASSS-----SANDAV-----SVNASIVKQGGDLVRDLKGGKASKPEIDAAYKTL 852  
 QY 158 ---AEWETKNADAIVKGAITELEKAYASDNQAI--LDSLGLKTSFDLLQTLALQSVAANN 212  
 DB 853 ELKAAQYKTLTGQDWKPGTVPPTAAPSASAPSVGGVNDVAQILS-----QITAGD 903  
 QY 213 KAELLKEMODNPVVPCKTPTAIAQSLVDQDTATQTEKGNAGICDAYFACONASGAVEN 272  
 DB 904 KVRRLKSAKADKATVDAAYKTLTSLKADYKAATGSD-WKPGTT-----APAPAAAPVK 956  
 QY 273 AKSN-----SISNIDSAAKAIATAKTAIAEAKKFPDPSPILOEAEQMVIOAEKDKNI-- 326  
 DB 957 KQKNPPDPASVLVNTLLNKIAQGGDKIRQLKSAKSEKSLVEAEVKULLALKUTDYKSLTG 1016



QY 327 ---KPADGSDVPNPGTT-----VGSKQOGSSIGSI----- 354  
Db 1017 QEWKP--GTVPAPTIVNVIDLTTG--DGSVDYGSVLKIQAGDKIRKLKSEKAAKNVI 1072  
QY 355 --RVSMLL-----DDAENETASILMSGFROMIHMFENTENPDSQAAQOELA 397  
Db 1073 DPEVKTLALKGEYKTLGKDWTPDAKSEPAVV-----KKEASPVSWASPAKDELTOEIN 1127  
QY 398 AQARAAKAGDSDSAAALADAQALEAL 426  
Db 1128 AQGEKVRAGKNGKAKEVIDAEVAKLLAL 1156

Search completed: February 7, 2002, 21:42:42  
Job time: 596 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:41:09 ; Search time 172 Seconds  
(without alignments)  
416.706 Million cell updates/sec

Title: US-09-391-606-8  
Perfect score: 2412  
Sequence: 1 MVNPIGPGIDETERTPPAD.....QKLISEEDLNSAVDHHHHH 490

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL\_17.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mmc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2214	91.8	715	2 Q46166	Q46166 chlamydia p
2	2188	90.7	651	2 Q927H7	Q927H7 chlamydia p
3	451.5	18.7	647	2 Q84627	Q84627 chlamydia t
4	195.5	8.1	1327	2 Q9X7M2	Q9X7M2 staphylococ
5	195.5	8.1	2478	2 Q9RL69	Q9RL69 staphylococ
6	193.5	8.0	2478	2 Q9LCH2	Q9LCH2 staphylococ
7	188	7.8	2055	2 Q85472	Q85472 abiectrophia
8	182.5	7.6	2481	2 Q99QR6	Q99QR6 staphylococ
9	182	7.5	1545	2 Q9RDO1	Q9RDO1 streptomyce
10	181	7.5	697	5 Q9NDJ0	Q9NDJ0 plasmodium
11	180	7.5	2271	2 Q99QY4	Q99QY4 staphylococ
12	178	7.4	1822	2 Q07290	Q07290 streptococ
13	174.5	7.2	1795	2 Q9LCJ9	Q9LCJ9 staphylococ
14	173	7.2	2016	5 Q9BXT0	Q9BXT0 plectreuris
15	171	7.1	627	1 Q9HS86	Q9HS86 halobacteri
16	171	7.1	993	2 Q99QZ5	Q99QZ5 staphylococ
17	170	7.0	1128	5 Q26947	Q26947 trypanosoma
18	168.5	7.0	1038	10 Q9AS09	Q9AS09 oryza sativ
19	166.5	6.9	2232	5 P91365	P91365 caenorhabdi

20	166.5	6.9	6713	2 Q99U54	Q99U54 staphylococ
21	163.5	6.8	956	2 Q9LON7	Q9LON7 streptomyce
22	163.5	6.8	2045	2 Q9AOK5	Q9AOK5 streptococ
23	161.5	6.7	3381	2 Q9KX33	Q9KX33 streptococ
24	160.5	6.7	344	1 P71409	P71409 halobacteri
25	160.5	6.7	1156	2 Q925A4	Q925A4 streptomyce
26	160.5	6.7	2178	2 Q9KWR3	Q9KWR3 streptococ
27	159	6.6	1579	11 Q99WP1	Q99WP1 mus musculu
28	159	6.6	1684	11 Q9WTQ5	Q9WTQ5 mus musculu
29	158	6.6	1365	2 Q49525	Q49525 mycoplasma
30	158	6.6	2186	2 Q99TB0	Q99TB0 staphylococ
31	157.5	6.5	1029	10 Q9LDB1	Q9LDB1 oryza sativ
32	157	6.5	1344	2 Q49545	Q49545 mycoplasma
33	157	6.5	6677	5 Q9N435	Q9N435 caenorhabdi
34	156.5	6.5	641	2 Q9KS57	Q9KS57 vibrio chol
35	156	6.5	845	5 Q9Y1P8	Q9Y1P8 plasmodium
36	156	6.5	990	13 Q91803	Q91803 xenopus lae
37	156	6.5	993	10 Q9ASL3	Q9ASL3 oryza sativ
38	155.5	6.4	582	5 Q16967	Q16967 aplysia cal
39	155.5	6.4	1041	10 Q9ASA4	Q9ASA4 oryza sativ
40	155.5	6.4	2273	5 Q9U141	Q9U141 leishmania
41	155	6.4	892	3 Q9P3P5	Q9P3P5 neospora
42	155	6.4	969	5 Q9ND19	Q9ND19 plasmodium
43	155	6.4	1147	2 Q87848	Q87848 streptomyce
44	155	6.4	1569	2 Q54183	Q54183 streptococ
45	154.5	6.4	1262	5 Q20684	Q20684 caenorhabdi

## ALIGNMENTS

RESULT 1	
Q46166	Q46166 PRELIMINARY; PRT; 715 AA.
AC	Q46166; 1996 (TREMREL. 01, Created)
DT	01-NOV-1996 (TREMREL. 01, Last sequence update)
DT	01-NOV-1996 (TREMREL. 01, Last sequence update)
DE	76 KDA PROTEIN.
OS	Chlamydia pneumoniae (Chlamydia pneumoniae)
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia
OX	NCBI_TaxID=83558;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=94156481; PubMed=7509320;
RX	Perez-Melgosa M., Kuo C.C., Campbell L.
RT	*Isolation and characterization of a gene encoding a Chlamydia
RT	pneumoniae 76-kilodalton protein containing a species-specific
RT	epitope.
RL	Infect. Immun. 62:880-886(1994).
DR	EMBL; L23921; AAA23117.1; -
SQ	SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;

Query Match	91.8%; Score 2214; DB 2; Length 715;
Best Local Similarity	98.9%; Pred. No. 1.8e-114;
Matches	454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY	1 MVNPIGPGIDETERTPPADLSAOGLEASAAANKSAEORIAAGAKPKESKTSVERWSI 60
DB	257 LVNPIGPGIDETERTPPADLSAOGLEASAAANKSAEORIAAGAKPKESKTSVERWSI 316
QY	61 LRSVAVNLSLADKLGIASSNSSSTSRSDVDSTTATATPTPPPTSDYKTAQATYDT 120
DB	317 LRSVAVNLSLADKLGIASSNSSSTSRSDVDSTTATATPTPPPTSDYKTAQATYDT 376
QY	121 IFTSTSLADIQAALVLSQDAVTNFKDAAETAEAEWETKNADAIKVAQITELAKY 180
DB	377 IFTSTSLADIQAALVLSQDAVTNFKDAAETAEAEWETKNADAIKVAQITELAKY 436
QY	181 ASDNQAILDLSLGLTSPDLLOALLQSVANNKKAELLKEMQDNVPVPGKTPATAQSLVD 240
DB	437 ASDNQAILDLSLGLTSPDLLOALLQSVANNKKAELLKEMQDNVPVPGKTPATAQSLVD 496



QY 241 QTDATATQIEKGNALGDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTQIAEAQ 300  
DB 497 QTDATATQIEKGNALGDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTQIAEAQ 556  
QY 301 KKFPPDSPILQEAQOMVIOAEKDLKNPKADGSDVNPNGTGVGSKQOGSSIGSIRVSMLL 360  
DB 557 KKFPPDSPILQEAQOMVIOAEKDLKNPKADGSDVNPNGTGVGSKQOGSSIGSIRVSMLL 616  
QY 361 DDAENETASILMSGFQOMIHMTENPDSOAAQOELAAQARAARAKAGDDSAALADAK 420  
DB 617 DDAENETASILMSGFQOMIHMTENPDSOAAQOELAAQARAARAKAGDDSAALADAK 676  
QY 421 ALAALGKAGQOQOILNALGQIASAAVVSAGVLPLOQVL 459  
DB 677 ALAALGKAGQOQOILNALGQIASAAVVSAGVLPLOQVL 715

## RESULT 2

Q927H7 ID Q927H7 PRELIMINARY; PRT; 651 AA.  
AC Q927H7;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE CHLPN 76 KDA HOMOLOG\_1 (CT622).  
GN CPN0728 OR CPJ0728 OR CP0018.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,  
RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CWL029 from USA".  
RL Nucleic Acids Res. 28:2311-2314(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,  
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
RA Bowman C., Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J.,  
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
RT pneumoniae AR39".  
RL Nucleic Acids Res. 28:1397-1406(2000).  
DR EMBL; AE001654; AAD18867.1;  
DR EMBL; AP002547; BAA98935.1;  
DR EMBL; AE002165; AAF37914.1;  
DR TIGR; CP0018;  
KW Complete proteome.  
SQ SEQUENCE 651 AA; 68217 MW; 47AE6C3FF2FF0123 CRC64;

Query Match 90.7%; Score 2188; DB 2; Length 651;  
Best Local Similarity 98.9%; Pred. No. 4.4e-113;  
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MWNPICGPIDETERTPPADLSAQGLEASAAKSAEAOQRIAGAEAKPKESKTDTSVERWSI 60  
|||||

DB 1 MWNPICGPIDETERTPPADLSAQGLEASAAKSAEAOQRIAGAEAKPKESKTDTSVERWSI 60  
QY 61 LRSAYNALMSLADKLGIASSNSSSTSRSDVDTATATPTPPPTSDYKTOAQATYDT 120  
DB 61 LRSAYNALMSLADKLGIASSNSSSTSRSDVDTATATPTPPPTSDYKTOAQATYDT 120  
QY 121 IFTSTSLADIOAALVSLQDVTNINIKDTAATDEETAIAAEWETKNADAIVKGAQITELAKY 180  
DB 121 IFTSTSLADIOAALVSLQDVTNINIKDTAATDEETAIAAEWETKNADAIVKGAQITELAKY 180  
QY 181 ASDNQAILDSLGLTSTFDLLQTFALLQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVD 240  
DB 181 ASDNQAILDSLGLTSTFDLLQTFALLQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVD 240  
QY 241 QTDATATQIEKGNALGDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTQIAEAQ 300  
DB 241 QTDATATQIEKGNALGDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTQIAEAQ 300  
QY 301 KKFPPDSPILQEAQOMVIOAEKDLKNPKADGSDVNPNGTGVGSKQOGSSIGSIRVSMLL 360  
DB 301 KKFPPDSPILQEAQOMVIOAEKDLKNPKADGSDVNPNGTGVGSKQOGSSIGSIRVSMLL 360  
QY 361 DDAENETASILMSGFQOMIHMTENPDSOAAQOELAAQARAARAKAGDDSAALADAK 420  
DB 361 DDAENETASILMSGFQOMIHMTENPDSOAAQOELAAQARAARAKAGDDSAALADAK 420  
QY 421 ALAALGKAGQOQOILNALGQIASAAVVSAGVLP 454  
DB 421 ALAALGKAGQOQOILNALGQIASAAVVSAGVLP 454

## RESULT 3

O84627 ID O84627 PRELIMINARY; PRT; 647 AA.  
AC O84627;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE CHLPN 76 KDA HOMOLOG.  
GN CT622.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UV-3/CX;  
RX MEDLINE=99000809; PubMed=97841136;  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.,  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis".  
RL Science 282:754-759(1998).  
DR EMBL; AE001333; AAC68226.1;  
KW Complete proteome.  
SQ SEQUENCE 647 AA; 68525 MW; C0D14C2D74473625 CRC64;

Query Match 18.7%; Score 451.5; DB 2; Length 647;  
Best Local Similarity 29.6%; Pred. No. 2.3e-17;  
Matches 138; Conservative 85; Mismatches 192; Indels 51; Gaps 14;

QY 2 VNPICGPIDETERTPPADLSAQGLEASAAKSAEAOQRIAGAEAKPKESKTDTSVERWSIL 61  
DB 15 MNPIINGQI-----ASNSETKESTKESEA-----SPSASSSSWSFL 52

QY 62 RSAYNALMSLADKLGIASSNSSSTSRSDVDTATATPTPPPTSDY---KTQATAY 118  
DB 53 SSAXHALISLRD--AILNKNSSPTDSLS-OLEASTSIS-TYTRVAARDYNAKSNFDTAK 108  
QY 119 DTFTSTSLADIOAALVSLQDVTNINIKDTAATDEETAIAAEWETKNADAIVKGAQITELAKY 178  
DB 109 SGLENATTLAEYETKMDLMAALQDMERLAKQKAEVTRIKEALQOEKQEVID---KLNLQV 165



Qy	4	PIGPPTDE-----	TERTPPADLSA	QGLEASAANKSA	-BAQRTAGAEAKPKESKTD	----	54
		: :   :   :   :	:   :   :   :	: :   :   :	: :   :		
Db	741	PLNPDTTNEEVAE	IERINAAKVS	-GVKAIEATTTA	QDLERVKNEEISKIENITD	STOT 798	
Qy	55	-----	VERWSITLRS	AVNALMS-----	-----	LADKLG-----	IASSNSSSSFS 87
		: : :   :   :	:   :   :	:   :   :	:   :   :		
Db	799	KMDAYNEVKQA	TARKAQNATVS	NATNEEVAEADA	ADRAAQKGLHDIQVVK	SKOEVA	DT 858
Qy	88	RSADVSTTA-TA	PTPPPTSDYK	TQATAYDT-----	IFTSTLADIOAALVSL	----	137
		:   :   :   :	:   :   :	: :   :	: :   :		
Db	859	KSKVLDKINA	IOAQVKPAAD	---TEVENAYNTR	KOE:QNSNASTTEEKQA	AYTDELTK 915	
Qy	138	-QDVAITN	KIDTAATD-----	EETALAEWETKN	ADAIKVGAOITELAKY	SDNQALDS 190	
		:   :   :   :	:   :   :	:   :   :	:   :   :		
Db	916	KQEARTNL-DA	ANTSNDVT	TAKDNSIAAIN	QVOAATTKKSDAK-AEIA	QAKASERT	ATIEA 973
Qy	191	LGLKIT-----	SFDLLQ	TALLQSVANNK	AAELLKMQDN	PVVPKGT-----	231
		: :   :   :	:   :   :	:   :	:   :		
Db	974	MNDSTTEEQ	QAARKDQV	AVVTANADIDNAA	-----	ANNVDNKNAT	TEATIAITPDA 102
Qy	232	---PATAQSL	VDOTD	TATATQIEK	DGN-----	---AIGDAYFAGON	ASCA 269



```
Db 838 KMDAYNEVKQAATARKQATVSNATNEEVAEADAADVAQAQKGLHDIQVVKSKQEVADT 897
Qy 88 RSADVDSSTA-TAPTPPPPTSDDYKYQAQPAYDT-----IFTSTSLADIQAALVSL---- 137
Db 898 KSKVLDRKINAIOQAKVKAAD---TEVENAYNTRKQEIQNSNASTTEKQAAAYTELDTK 954
Qy 138 -QDAYTNIKDTAATD-----EETAIAAEWETKNADAIVKGAQITELAKYASDNOAILDS 190
Db 955 KQEARTNL-DAANTNSDVTTAKDNTSIAAINOVQAATTKKSDAK-AEIAQKASERKTAIEA 1012
Qy 191 LGKLT-----SFDLIQTALLOSVAANNKAAELLKEMQDNPPVPGKT----- 231
Db 1013 MNDSTTEEQQAQKDVQAVVTANADIDNAA-----ANNVDNAKTTEATIAATTPDA 1066
Qy 232 ---PAIAQSLVDQDTATATQIEKDN-----AIGDAYFAGQNASGA 269
Db 1067 NVKPAQAQAIADKVOAQETAI--DGNNGSTTEEKAQAQVQTEKTITADAADAHTNAE 1124
Qy 270 VENAKSNNSISNI-----DSAKAAIAT---AKTOIAEAKKFPDPSPILOEA--E 313
Db 1125 VEAAK-KAAIAKTEAIOQATTTKDNKEAIAATKANERKTAIAQODITAEIAAANDVD 1183
Qy 314 QMWIOAEKDLKNIKPADG-SDVPNPCTTVGSKQGGSSIGSIRVSMILLDDAENETASILM 372
Db 1184 NAVTOAN---SNIEAANSQNDVQAKTTGENSIDQVTPVNNKAT-----ARNEITAILN 1235
Qy 373 SGFRQIMHFNTEPNDSQAQOELAAQARAAGAAGDSAAA-----ALADAQKALEAA 425
Db 1236 NKLEIQATPDATDEEKQAADAE--ANTENGKANAQISAATTNAQVDEKANAEEAIAINAV 1293
Qy 426 LGKAGQOQGLNALGOITASAAVVSAGVLPLOQVLRIRARYQAVVEOKLISEEDLNSAV 483
Db 1294 TPKVVVKQAQKDEIDQLOAT-----QTNVINNDQNAATTEKEAAIQOLATAV 1340
```

```
RESULT 6
Q9LCH2 PRELIMINARY; PRT; 2478 AA.
AC Q9LCH2;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE FMTB.
GN FMTB.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RA Komatsuzawa H.;
RT "Staphylococcus aureus gene for affecting the methicillin
  resistance.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB025716; BAA93438.1;
SQ SEQUENCE 2478 AA; 262993 MW; 1C118EBE0DB03B34 CRC64;
```

```
Query Match 8.0%; Score 193.5; DB 2; Length 2478;
Best Local Similarity 23.6%; Pred. No. 0.02;
Matches 141; Conservative 90; Mismatches 212; Indels 155; Gaps 28;

Qy 4 PICPGPIDE-----TERTPPADLSAQGLEASAANKSA-EAORIAAGAAKPKESKTS--- 54
Db 780 PLNPDTTNEEVAERINAAKVS--GVKAIEATTTAQDLERVKKEISKIENIDSTQ 837
Qy 55 -----VERWSTLSAVNALMS-----LADKLG-----IASSNSSSTS 87
Db 838 KMDAYNEVKQAATARKQAQNASATNEEVAEADAADVAQAQKGLHDIQVVKSKQEVADT 897
Qy 88 RSADVDSSTA-TAPTPPPPTSDDYKYQAQPAYDT-----IFTSTSLADIQAALVSL--- 137
```

```
Db 898 KSKVLDRKINAIOQAKVKAAD---TEVENAYNTRKQEIQNSNASTTEKQAAAYTELDTK 954
Qy 138 -QDAYTNIKDTAATD-----EETAIAAEWETKNADAIVKGAQITELAKYASDNOAILDS 190
Db 955 KQEARTNL-DAANTNSDVTTAKDNTSIAAINOVQAATTKKSDAK-AEIAQKASERKTAIEA 1012
Qy 191 LGKLT-----SFDLIQTALLOSVAANNKAAELLKEMQDNPPVPGKT----- 231
Db 1013 MNDSTTEEQQAQKDVQAVVTANADIDNAA-----ANNVDNAKTTEATIAATTPDA 1066
Qy 232 ---PAIAQSLVDQDTATATQIEKDN-----AIGDAYFAGQNASGA 269
Db 1067 NVKPAQAQAIADKVOAQETAI--DGNNGSTTEEKAQAQVQTEKTITADAADAHTNAE 1124
Qy 270 VENAKSNNSISNI-----DSAKAAIAT---AKTOIAEAKKFPDPSPILOEA--E 313
Db 1125 VEAAK-KAAIAKTEAIOQATTTKDNKEAIAATKANERKTAIAQODITAEIAAANDVD 1183
Qy 314 QMWIOAEKDLKNIKPADG-SDVPNPCTTVGSKQGGSSIGSIRVSMILLDDAENETASILM 372
Db 1184 NAVTOAN---SNIEAANSQNDVQAKTTGENSIDQVTPVNNKAT-----ARNEITAILN 1235
Qy 373 SGFRQIMHFNTEPNDSQAQOELAAQARAAGAAGDSAAA-----ALADAQKALEAA 425
Db 1236 NKLEIQATPDATDEEKQAADAE--ANTENGKANAQISAATTNAQVDEKANAEEAIAINAV 1293
Qy 426 LGKAGQOQGLNALGOITASAAVVSAGVLPLOQVLRIRARYQAVVEOKLISEEDLNSAV 483
Db 1294 TPKVVVKQAQKDEIDQLOAT-----QTNVINNDQNAATTEKEAAIQOLATAV 1340
```

```
RESULT 7
O85472 PRELIMINARY; PRT; 2055 AA.
AC O85472;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
DE EXTRACELLULAR MATRIX BINDING PROTEIN (FRAGMENT).
GN EMB.
OS Abiotrophia defectiva.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Abiotrophia.
OX NCBI_TaxID=46125;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NVS-47;
RA Manganello R., van de Rijn I.;
RT "Cloning and Characterization of emb, a Gene Encoding the Major
  Adhesin of Streptococcus defectivus.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF067776; AAD03320.1;
FT NON_TER 2055
SQ SEQUENCE 2055 AA; 215640 MW; 9699C11DDE93E2FD CRC64;
```

```
Query Match 7.8%; Score 188; DB 2; Length 2055;
Best Local Similarity 23.9%; Pred. No. 0.032;
Matches 124; Conservative 68; Mismatches 197; Indels 130; Gaps 24;
```

```
Qy 19 ADLSAQGLEASAANKSAEQAQRIAGAEA-----KPKESKTSVERWSILRSVAVNALNSLAD 73
Db 1602 AKKADAIADAATSNADVTQAQDAGKNAINAVPQTPTAKTD-----AKNAVTAQAD 1651
Qy 74 KLGIASSNSSSTSRSDVDSTTATATPPPTSDDYKYQAQPAYDTIFTSTSLADIQA- 132
Db 1652 AKKDAIENDANILTREKDAAKAKVDAAE-----TKAKNAID-----AATSNADVTAK 1698
Qy 133 -----ALVSLQDAVT-----NIKDTAATDEETAI-----AAEWETKNADAIVKGAQITE 176
Db 1699 QNEGTKAINDVPQTPTAKTDKAKNAVDQAATDKKSAIENDPALTREKDAAKAKVDAAEATK 1758
```



QY 177 LAKYASDNQAIALSLGKLTSEFDLLOLQALLOSVANNKAAELLKEMQNPV--VPGKTPAI 234  
 Db 1759 -AKNAID-----AATSNADVTQAQKDAKGNAINAVP-QTPA 1792  
 QY 235 ---AQLSDQ-----TDATATQIEKD-GNAIGDAYFAGNAGAVENAKNSNI- 279  
 Db 1793 KTDAKNAVDAQATDKKAAIENDPALTREKDAKAAKAVDA--EAKKAKDAIDAATSNADVT 1850  
 QY 280 SNIDSAKAAI-----ATAKTOIAEA-----OKRFPDPSILOEAFOMVIOAEKDLN 325  
 Db 1851 AOKDAGKDAINAVPQTPAKTDKNAVDAQATDKKSAIENDPALTREKDAKAVKVADEA 1910  
 QY 326 IKPADGSDVPNPGTGVGSKOQS-SIGSI-RVSMILLDDABNETASILMSGFROMIHFN 383  
 Db 1911 KKAKDAIDAATSNADVTAKOTEGTQAINAVPQTPAKTDKNAVDAQATDKKNAIENDPA 1970  
 QY 384 TENPDQAQOELAAQARAAGDSDSAAALADAQAKALEALGKAGQOQG--TINALGQ 441  
 Db 1971 LTREKDAKAAKVADEAKKAKDAID--AATSNADV-----TAKONEGTKAINDVPQ 2019  
 QY 442 IASAAVVSAGVLPLOQVWIRARYQAVVEQKLISEEDLN 480  
 Db 2020 TPTAKTDKNAVQ-----QAATDKKSAIENDAN 2047

RESULT 8

Q99QR6 PRELIMINARY; PRT; 2481 AA.  
 ID Q99QR6 AC Q99QR6  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DE FMTB PROTEIN.  
 GN FMTB(MRP) OR SAI1964.  
 OS Staphylococcus aureus subsp. aureus N315.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Staphylococcus.  
 OX NCBI\_TaxID=158879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,  
 RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Oi Y.,  
 RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,  
 RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,  
 RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,  
 RA Ogasawara N., Hayashi H., Hiramatsu K.  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus".  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL; AP003136; BAB43253.1;  
 KW Complete proteome.  
 SQ SEQUENCE 2481 AA; 263767 MW; ELEAB99B81665E8 CRC64;

Query Match 7.68; Score 182.5; DB 2; Length 2481;  
 Best Local Similarity 23.18; Pred. No. 0.081;  
 Matches 136; Conservative 96; Mismatches 219; Indels 139; Gaps 26;  
 QY 4 PIGPCTDE-----TERPPADLSAQLGSAANKSA-EAORIAGAEAKPKESKTDG--- 54  
 Db 763 PLNPDTEEEVAERINAAKVS--GVKAIEATTTADLERVNEEFKIENTIDSTQT 820  
 QY 55 -----VERMSILRSVAVNALMSLADKLGITASSNSSSTSRSD-----VDSTTATAPT 101  
 Db 821 KMDAYKEVROAATARKAQNATVSNATDEEVAEANAADVAQTEGLHDIFQVVKSOQEVADT 880  
 QY 102 PP-----PPTSDDYKTKQATAYDT-----IFTSTSLADQAALVSL--- 137  
 Db 881 KAKVLDKINATQTAQVKPAD---TEVENAYNTRKQIIONSNASTTEKEEAAYTELDAK 937  
 QY 138 -QDAVTNIKDPAATD-----EETAIAAEWETKNADAIVKCAQITELAKYASDNQAILDS 190  
 Db 138 -QDAVTNIKDPAATD-----EETAIAAEWETKNADAIVKCAQITELAKYASDNQAILDS 190

Db 938 KQEARTNL-DAANTNSDVTTAKDNGIAAINOVQAATTKKSDAK-AEIAQKASERKTAIEA 995  
 QY 191 LGKLT-----SFDLLOLQALLOSVANNKA-----AELLKEMQNPVVPKGT--PAI 234  
 Db 996 MNDSTTEBQQAOKDKVQAVVTANADIDNATANTDNDNAKTNTNEATTAATIPDANVKPAA 1055  
 QY 235 AQLSDQTDATATQIEKDGA-----IGDAYFAGNAGAVENAKNSNN 277  
 Db 1056 KQAIADKVAQETAIDANNSTTEKEAKAQQOVOTEKTAADAAIDAASHNVVEEAAK-NA 1114  
 QY 278 SISI-----DSAKAIAIAT-----AKTOIAEAKKKFPDPSILOEA--EOMVIOAEK 321  
 Db 1115 EIAKIEAIQPATTTKDKNAQKAIATKANERKTAIAQTODITAEIEAANADVDNAVTOAN- 1173  
 QY 322 DLKNIKPADG-SDVPNPGTGVGSKOQSSIGSIRVSMILLDDABNETASILMSGFROMIH 380  
 Db 1174 --SIEAANSNDVDAQTTGETSIDQVTPVNNKAT-----ARNETAILNNKLQELQA 1226  
 QY 381 MFNTENPDQAQOELAAQARAAGDSDSAAA-----ALADAQKALEALGKAGQOQ 433  
 Db 1227 TPDATDEKQAADAE--ANTENKANQAISAATTTNAQVDEAKANAIAINAVTPKVVKKQ 1284  
 QY 434 GILNALGOIASAAVVSAGVLPLOQVWIRARYQAVVEQKLISEEDLSAV 483  
 Db 1285 AAKDEIDQLQAT-----QTNVINNDQATNEEKEAAIQOLATAV 1323

RESULT 9

Q9RDQ1 PRELIMINARY; PRT; 1545 AA.  
 ID Q9RDQ1 AC Q9RDQ1  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE PUTATIVE SECRETED PROTEIN.  
 GN SC4A7.11.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RC Seeger K.J., Harris D.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RC Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RC MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; AL133423; CAB62715.1;  
 DR InterPro; IPR000772; Ricin\_B\_lectin.  
 DR Pfam; PF00652; Ricin\_B\_lectin; 1.  
 DR SMART; SM00458; RICIN; 1.  
 SQ SEQUENCE 1545 AA; 161519 MW; 81EF325143593AEA CRC64;

Query Match 7.58; Score 182; DB 2; Length 1545;  
 Best Local Similarity 24.6%; Pred. No. 0.048;  
 Matches 130; Conservative 67; Mismatches 205; Indels 126; Gaps 24;  
 QY 11 DETERTPPADLSA--QGLEASAANKSAEAOIAGAEAKPK---ESKTDSEVERWS----- 59  
 Db 741 DAKGRAETADTTATEKRGAEAKDAKAKALRDDAWDAEQKANAARAKADAKEAYASAS 800



QY 60 -----ILRSVNMALSLADKLGIA-----SNSSSTSRSDVDTATPTPPPT---S 107  
Db 801 EDNAQEAARKAANDADAADAETAANAARSEADKATKAAADADAATAAEAAAKRGRADS 860  
QY 108 DDYKTOAQATVDTFTSTSLADICAAALVSLQDAVTKDKTAADEETAI-----AA 158  
Db 861 DAAQAAMLTADAARVATSA-----ADAIKASK-TAATAARTAVELADDAEQHAA 910  
QY 159 EWETKNADAIVKGAQITELAKYASDNOAILDSLGKITSFDLLQTLALLOSANNKAEALL 218  
Db 911 D-AKKEADAARAE-VTALA-----GANESTGYAVTTAAQAAVADAGNSAAQVA 955  
QY 219 KEMQD-----NPVPGKTPAIAQSLVDOTATATQIEKDNAGNIGDFAFQONASGAVENA 273  
Db 956 APANDAIOQLGSPYV---TTDSAGLAVLTGOSKTTIAEQQQQVAEAEH-----AQNAESA 1007  
QY 274 KSNNSINIDS--AKAAIATAKTOI---AEAQKFPD-----SPILQEAQEWVIOA 319  
Db 1008 AQAOSVANASGDSKAAATLAAEALGYAADARNSAKEALGYSAAEASVATQAAQSLRTTI 1067  
QY 320 EKDLNITKPADGSDVPNPGTTVGGSKOOGSIGIRVMSLDDAENETASILMSGFQMI 379  
Db 1068 AYDTQATKDAAAD-----SAG--RAEGHAEDARDSADAAL-----1103  
QY 380 HMFENTPDSAAQ-----QELAAQAR-AKAAGDDSA--ALADAQKALEAALCKAG 430  
Db 1104 -----DAEARSADTAEQAEADARDADHAATGAARAEAAKQAKVAESAQAQAAE 1155  
QY 431 QOQGLNALGQIAAAVVSAGVLPLOQVLMIRARYQAYVEOKLISEED 478  
Db 1156 QAEREANA-EQIDKGTVDQGTGAPIGDVYVVDHIEKIGEPVVKQSD 1202

RESULT 10  
Q9NDJ0 PRELIMINARY; PRT; 697 AA.  
AC Q9NDJ0;  
AC Q9NDJ0;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE MEROZOITE SURFACE PROTEIN 3B.  
GN MSP-3B.  
OS Plasmodium vivax (strain Belem).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=31273;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BELEM;  
RA Gallinski M.R., Al-Khedery B., Ingravallo P., Corredor-Medina C.,  
RA Barnwell J.W.;  
RT "Plasmodium vivax merozoite surface proteins-3 beta and -3 gamma share  
structural similarities with plasmodium vivax MSP-3 alpha and define a  
new gene family in Plasmodium.";  
RT submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AF099662; AAF78287.1;  
DR InterPro; IPR000122; ChemoTaxis\_transducer.  
SQ SEQUENCE 697 AA; 75685 MW; 0AF2AE9801A956A7 CRC64;

Query Match 7.5%; Score 181; DB 5; Length 697;  
Best Local Similarity 20.8%; Pred. No. 0.02;  
Matches 124; Conservative 81; Mismatches 200; Indels 192; Gaps 21;

QY 1 MWNPTGPGIDETERTPPA-----DLSAQGLEASAANKSAEQRI 40  
Db 53 IVPNGSDVDTEGDGEALGQNGSPKSAEPKVHAEVVKESLKSAQNAKAEAKA 112  
QY 41 AGAEAKPKESKTSVVERSI-----LRSVAVN 66  
Db 113 AKAAESAKNTLDLEKVNVPTELNVKKFAESAATEAQKQENIATEAEKKVAEANGEVV 172  
QY 67 ALMSLADKLGIASSNS-----SSSTSRSDVDTTAA-----97

Db 173 ELQKLKDEVDKAAKAKQOLQKAEIAEHAHVAKQAVKTEAEKAKQDATTAKAYAIKETGTS 232  
QY 98 -----TATPPPPPTSDVKTQAOATAYDIFTSTSLA-----DIQAALVSLQDAVT 142  
Db 233 KSENVTKAIDMAKKEEETKNOASIASENADRAAKAAQAEVKEIKDENKEISOLENEIT 292  
QY 143 NIKDTAATDEETAIAAEWETKNADAIVGQAI-----TELAKYASDNOAILDSLGKLTSS 196  
Db 293 KLGDIINTVKELASNAEDASKNAKKEWKAQIAAEVAKAEKAKIEAENANFLAEKAK--- 349  
QY 197 FDLQTLALLOSANNKAAELKEMODNPVPGKTPAIAQSLVDOTATATQIEKDNAGNAI 256  
Db 350 ----QTA--EKIAKTSKSTEKITE-----EVRKATEFAKTAGDATQTAATE-----AA 391  
QY 257 GDYFAGQNASGAVENAKSNNSINSDSAKALATATAKTO---IAEAQKFPDPSILQEE-- 311  
Db 392 GDSVSEQKQKVLESIK--OKAESALQASKDAI--KARTEAENFLEIAKEVPRKAAKEEA 449  
QY 312 -----AEQMVIOAEKDLKNIKPADGSDVPNPGTTVGGSKOOGSIGIRVMSLDDAEN 365  
Db 450 QKAATSADAETKALTAIEEVNKSDAS-----EN 478  
QY 366 ETASILMSGFQMIHMFNTE-NDPSQAQOELAAQARAARAKAAGDSSAAALADAQKALEA 424  
Db 479 EKKKI-----ETEANATAGEAQAARAFKAFAEADAADAKDTNEAVTLAVAKEKVEK 526  
QY 425 ALGKAGQOQGLNALGQIAAAVVSAA---GVLPLOQVLMIRARYQAYVEOKLISEE 477  
Db 527 AL-KAAKEAKKAN---EKASYALIRTKQYALEPLE---ITSEAGYNITEKEEQVKEE 577

RESULT 11  
Q99QY4 PRELIMINARY; PRT; 2271 AA.  
AC Q99QY4;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SA2447 PROTEIN.  
GN SA2447.  
OS Staphylococcus aureus subsp. aureus N315.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=158879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,  
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,  
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,  
RA Hiramatsu H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,  
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,  
RA Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus.";  
RT Lancet 357:1225-1240(2001).  
RL EMBL; AP003138; BAB43752.1;  
DR Complete proteome.  
KW SEQUENCE 2271 AA; 227847 MW; 7C2A7040D6C8289D CRC64;

Query Match 7.5%; Score 180; DB 2; Length 2271;  
Best Local Similarity 18.9%; Pred. No. 0.1;  
Matches 79; Conservative 96; Mismatches 187; Indels 56; Gaps 12;

QY 22 SAQGLEASAANKSAEQRIAGAEAKPKESKTSVERWSILRSAYNALMSLADKLGIASSN 81  
Db 1516 SASASESSSTSDSTSDTSASMQSESQTSASLSLSLSTSTSNRMT-----IASLS 1570  
QY 82 SSSSTSRSDVDTATPTPPPTSDDYKTOAOT-AYDTFTSTSLADIQAALVSLQDA 140  
Db 1571 TSVSTSGSTSESTSESDSTSLSDSQTSRSTSRSTASGSASTSTSTSDSRSTASTS-- 1628







QY 323 LKNIKPADG--SDVPNPGTTVGSKQGGSSIGSIRVSMILLDDAENETASILMSGFROMIHM 381  
 Db 1194 -SNIEANSQNDVQAKTTGTTSIDQVTPVTKKAT-----ARNETAILNKNKLOEIQT 1247  
 QY 382 FNTENPDSQAQOCEL-----AAQARAARAAAGDSAAAALADAKALEAALGKAGQOQGI 435  
 Db 1248 PDATDEKQADAEANTENGKALQATAA-ATTNAQVDEAKTNAEAAINAVTPKVVKKQTA 1306  
 QY 436 LNALGQTASAAVVSAGVLPQQVLWIRARYQAYVEQKLISEEDLNSAV 483  
 Db 1307 KDEIDQLOAT-----QTNWINNDQNTATTEKEAAIQOLATAV 1343  
 RESULT 14  
 Q9BIT0 PRELIMINARY; PRT: 2016 AA.  
 AC Q9BIT0;  
 DT 01-JUN-2001 (TREMELrel. 17, Created)  
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
 DE FIBROIN 3 (FRAGMENT).  
 OS Plectreureys tristis (Spider).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Araneomorphae; Haplogynae; Plectreureidae; Plectreureys.  
 OC NCBI\_TaxID=33319;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21179804; PubMed=11283372;  
 RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;  
 RT "Extreme Diversity, Conservation, and Convergence of Spider Silk  
 Fibrin Sequences";  
 RL Science 291:2603-2605(2001).  
 DR EMBL; AF350283; AAK30612.1;  
 FT NON\_TER  
 SQ SEQUENCE 2016 AA; 196162 MW; F53A47DB185826F1 CRC64;  
 Query Match 7.2%; Score 173; DB 5; Length 2016;  
 Best Local Similarity 21.5%; Pred. No. 0.21;  
 Matches 103; Conservative 83; Mismatches 212; Indels 88; Gaps 16;  
 QY 19 ADLSAQGLEASAANKSAEAOIRAGAEAKPKESKTDVSVERMSILRSVAVNMLSLADKLGIA 78  
 Db 1241 ASASAYELEFSDSSSSAVAVASSQGSYDTSDD-----FSSASSAAAAAASAYESFLD 1296  
 QY 79 SSSSSSTSRADVDSTATATPTPPPTSDYKTAQATAYDTIPTSTSLADIAALVSLQ 138  
 Db 1297 ASSSSSSAAAASSQSSYETSSDLVSASSAAAAAASASAYSQFLDASSSSNAATTSR 1356  
 QY 139 ----DAVTNIRKTDATDETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDSLGL 194  
 Db 1357 QSSYDTSDFSSASIAAAAAAASAYESQFSDA-----SSSSNAAAAASSQSSYD----- 1407  
 QY 195 TSFDLLOTA-----LLQSVANNKRAELKEMQNPVPGKTPAIAQSLVD-QTDAT 245  
 Db 1408 TSSDLVSAASAYESQFLDASSSSNA-----TTSSQSSYDTSDFS 1452  
 QY 246 ATQIEKQCNATGDAY-----FAGQNASGAVENAKSNNSISN-IDSAKAAIATAKTQI 296  
 Db 1453 SASIAAAAAAASSTESQFSDASSSSNAAAAAAASQSSYDTSDDLVSASSAAAAAASASS 1512  
 QY 297 AEAKKFPDSPILQEA-----EQMVIOAEKDLKNIKPADG-----SDVPNPG 338  
 Db 1513 YESQ--FSDASSSSNAAAAAAASQSSYDTSDDLVSASSAAAAAASAYESQFSDASSSR 1570  
 QY 339 TTVGSKQGGSSIGSIRVSMILLDDAENETASILMSGFR-QMIMFENTENPDSQAQOE-- 395  
 Db 1571 NAAAAAASQSSYDTP--SSDLVSASSAAAAAASAYESQFLDASSSSNAATTSQSS 1628  
 QY 396 -----LAAQARAARAAAGDSAAAALADAKALEAALGKAGQOQ-----ILNALGOI 442  
 Db 1629 YDTSDFSSASIAAAVAASAYESQFSDASSSSKAAAAAASQSSYDTSDDLVSASSAA 1688

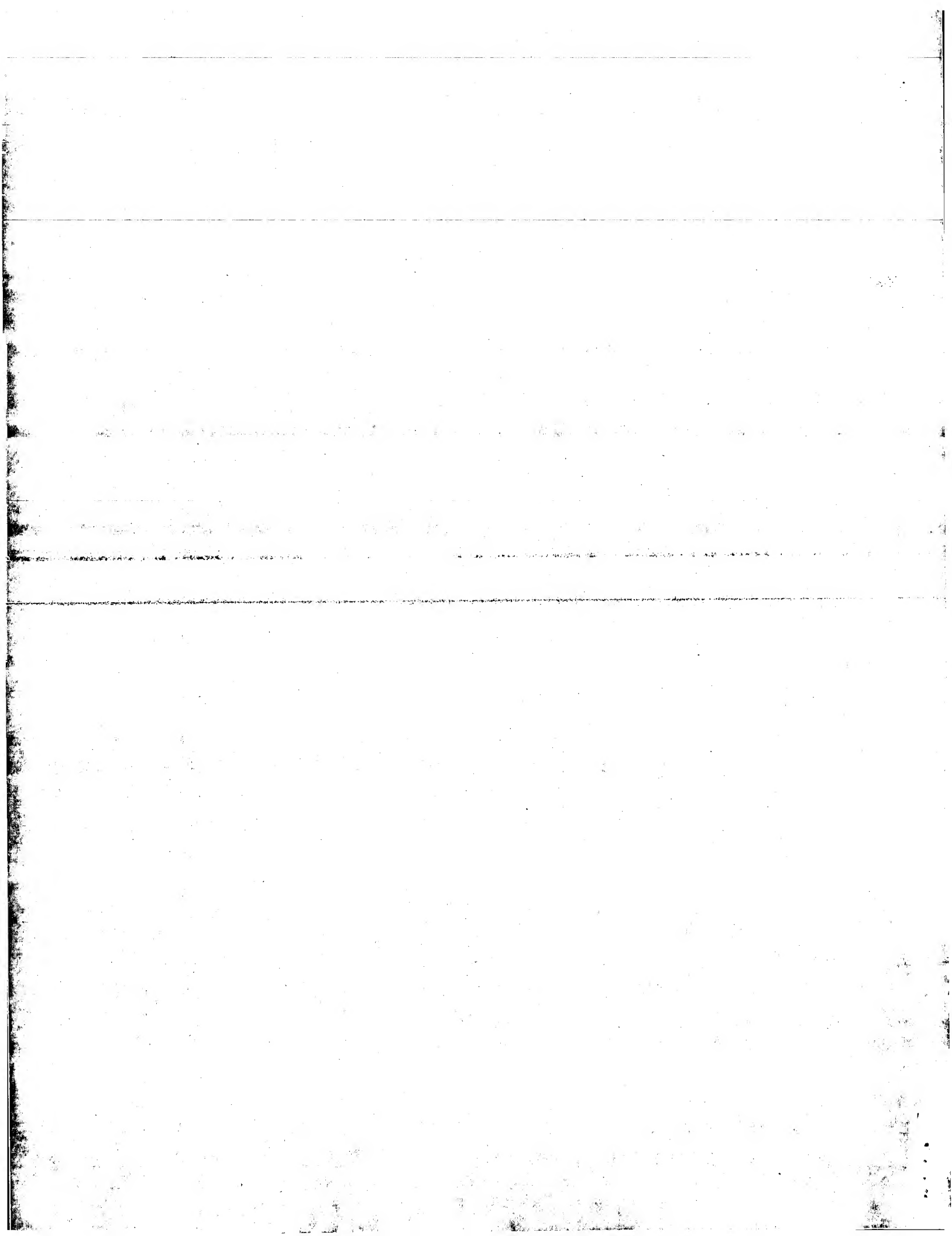
QY 443 ASAAVVSA 450  
 Db 1689 AAAASASS 1696  
 RESULT 15  
 Q9HS86 PRELIMINARY; PRT: 627 AA.  
 ID Q9HS86  
 AC Q9HS86;  
 DT 01-MAR-2001 (TREMELrel. 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)  
 DE HTR14 TRANSDUCER.  
 GN HTR14 OR VNG0355G  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 OC NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogha J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL; AE004994; AAG18922.1;  
 DR InterPro; IPR000122; Chemotaxis\_transducer.  
 DR InterPro; IPR000658; DUF5.  
 DR InterPro; IPR003660; HAMP.  
 DR Pfam; PF00672; DUF5; 1.  
 DR Pfam; PF00015; MCPsignal; 1.  
 DR PRINTS; PR00260; CHEMTRNSDUCR.  
 DR SMART; SM00304; HAMP; 2.  
 DR SMART; SM00283; MA; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 627 AA; 65655 MW; E5F2867FA6CA2A75 CRC64;  
 Query Match 7.1%; Score 171; DB 1; Length 627;  
 Best Local Similarity 23.4%; Pred. No. 0.063;  
 Matches 120; Conservative 87; Mismatches 184; Indels 122; Gaps 27;  
 QY 17 PPADLSAQGLEASAANKSAEAOIRAGAEAKPK-----ESKTDVSVERMS-ILRSVAVNMLSL 71  
 Db 145 PALDESYPGAFGESITEMADSLAEYTAELDKTAELHQOAELEROSEQUALVLDLSEA 204  
 QY 72 ADKLGIASSSSSTSRSDVDSTATATPTPPPTSDYKTAQ-----QTAIDYDTFTS 124  
 Db 205 TD-----AARAGDL---TATVDAALDVTDDHRAAVDFDQLLETADTISDI 249  
 QY 125 TSLADIAQAL-----VSLQDAVTNIKDTA--ATDEETPAIAAEWETKNADA 167  
 Db 250 QSFSDAVLAVSRITDERVDAVADRSAAVSESVTEIADGANQOTNOLNIAAEMDVTSATV 309  
 QY 168 IKVGAQITELAKYASDNQAIL-----DSLGLK-TSFDLLOTLQSVANNKAAELLKEMQ 222  
 Db 310 EEIAAASADVAKTA---QAAADRGDEGEVEETIEALRALREQSQAETVESLAEEVE 366  
 QY 223 DNPVPGKTPAIAQSLVDOTDATA--TQIE-----KQNGAIGDAYFAGQNASGAVENAKS 275  
 Db 367 R---IDGIT-ALIEDIAEETNMLNALSIEAARTGSDG--GFAVVADEVKDLAETREQ 420  
 QY 276 NNSTNSIDSA-----KAATATAKTQIAEAKKFPDSP-ILQEAQMVIOAEKDLKNKP 328  
 Db 421 AADISEIVDAVTEKAEDASIAIGEVD-AEVEKTKITKAEGLVRDFEIV----DEVANVNH 475  
 QY 329 A--DGSVDPNPG-----ITVGGSKQGGSSI-----GSIRVSMILLDD 362



Db	476	AVQETSDATQGAQSVTDVVGMEVEVASSEETAESDTVADNAAEQTDATDEVADQMD	535
Qy	363	AENETAST--LMSGFROMIHMFNTENPDSQAAQOELAAQAPAAK--AAGDDSRAAALADA	418
Db	536	LAEQTAALAGMLDDF-----TVPADAGTADQSVADDSPTAQPPAADDEPAAAVVDQ	587
Qy	419	QKALEAALGRAGQOQOQILNALGQIASAAVVSAG	451
Db	588	QPASDAE-----DEEGVPDSGGE--SVAVSDGG	613

Search completed: February 7, 2002, 21:41:15  
Job time: 4644 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:34:36 ; Search time 137.02 Seconds  
(without alignments)  
250.299 Million cell updates/sec

Title: US-09-391-606-9

Perfect score: 2261

Sequence: 1 MNPVIGPGIDETPPAD.....SAAVVSAGVLPQQVLWIRA 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_ll01.\*
- 1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.\*
  - 2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.\*
  - 3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.\*
  - 4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.\*
  - 5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.\*
  - 6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.\*
  - 7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.\*
  - 8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.\*
  - 9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.\*
  - 10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.\*
  - 11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.\*
  - 12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.\*
  - 13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.\*
  - 14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.\*
  - 15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.\*
  - 16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.\*
  - 17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.\*
  - 18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.\*
  - 19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.\*
  - 20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.\*
  - 21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.\*
  - 22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2241	99.1	490	21	AA1957
2	2188	96.8	651	21	AA1954
3	2184	96.6	452	21	AA1956
4	2162	95.6	478	20	AA1955
5	1850	81.8	583	21	AA1955
6	281.5	12.5	350	20	AA1957
7	231.5	10.2	361	21	AA13695
8	231.5	10.2	361	22	AA13695
9	181.5	8.0	1822	13	AA1955
10	164.5	7.3	331	20	AA1957
11	153.5	6.8	1237	21	AA1954

12	153	6.8	600	22	AA1954	C glutamic prote
13	152	6.7	1561	17	AA1954	S. mutans antigen
14	150.5	6.7	643	16	AA1954	Trypanosoma cruzi
15	148	6.5	729	22	AA1954	Mycobacterium tube
16	148	6.5	881	22	AA1954	S. cerevisiae apopt
17	147.5	6.5	1095	22	AA1954	S. epidermidis ope
18	146.5	6.5	571	21	AA1954	Arabidopsis thalia
19	143	6.3	1566	16	AA1954	Immunodominant ant
20	142	6.3	1607	19	AA1954	Mouse laminin Gl c
21	140.5	6.2	564	16	AA1954	Trypanosoma cruzi
22	140	6.2	876	22	AA1954	C glutamic prote
23	139	6.1	5024	22	AA1954	S. epidermidis ope
24	138	6.1	1576	21	AA1954	Human laminin 2 ma
25	138	6.1	1576	21	AA1954	Human laminin 2 po
26	138	6.1	1584	21	AA1954	Human laminin 2 ga
27	138	6.1	1609	19	AA1954	Human laminin Gl c
28	138	6.1	1609	21	AA1954	Human laminin 2 ga
29	138	6.1	1609	21	AA1954	Human laminin 8 po
30	138	6.1	1617	21	AA1954	Human laminin 2 ga
31	138	6.1	2742	21	AA1954	Human APC protein
32	138	6.1	2842	15	AA1954	Adenomatous polypo
33	138	6.1	2843	13	AA1954	APC gene product 1
34	138	6.1	2843	13	AA1954	Adenomatous polypo
35	138	6.1	2843	16	AA1954	Adenomatous polypo
36	138	6.1	2843	18	AA1954	Human adenomatous
37	138	6.1	2843	19	AA1954	Human APC protein
38	138	6.1	2843	19	AA1954	Human APC protein
39	138	6.1	2843	19	AA1954	Human APC protein
40	138	6.1	2843	21	AA1954	Human APC protein
41	138	6.1	2860	15	AA1954	Adenomatous polypo
42	138	6.1	2973	19	AA1954	Human APC protein
43	138	6.1	2973	21	AA1954	Human APC protein
44	138	6.1	2973	22	AA1954	Transcriptional ac
45	137.5	6.1	2453	21	AA1954	HNRCP protein sequ

## ALIGNMENTS

RESULT 1

AA1957  
ID AA1957 standard; Protein; 490 AA.

XX AA1957;

AC AA1957;

DT 26-MAR-2001 (first entry)

XX

DE C. pneumoniae 76 kba protein truncation mutant fusion protein.

XX 76 kba protein; bactericidal; diagnosis; prevention;

KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;

KW acute respiratory disease; cough; sore throat; hoarseness; fever;

KW vaccine; immunisation; treatment; fusion protein; truncation mutant;

KW mtein.

XX Chlamydia pneumoniae.

OS Synthetic.

XX Key

FT Region

FT Location/Qualifiers

FT 453..490

FT /note= "This part of the sequence is unrelated to the C. pneumoniae 76 kba protein"

FT WO2000066739-A2.

PD 09-NOV-2000.

XX 03-MAY-2000; 2000WO-CA00511.

XX 03-MAY-1999; 99US-0132270.

XX 30-JUN-1999; 99US-0141276.

XX (AVET ) AVENTIS PASTEUR LTD.



XX Murdin AD, Oomen RP, Wang J, Dunn P;  
 XX WPI: 2000-687542/67.  
 DR N-PSDB; AAD02066.  
 XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,  
 PT useful for vaccinating against Chlamydia infections -  
 XX Claim 33; Fig 3; 112pp; English.  
 PS  
 XX The present sequence is a fusion protein comprising a  
 CC truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded  
 CC residues. C. pneumoniae 76 kDa protein is used in the  
 CC diagnosis, prevention and treatment of C. pneumoniae  
 CC infections (e.g. pneumonia, upper respiratory  
 CC tract disease, bronchitis, sinusitis and acute respiratory  
 CC disease such as cough, sore throat, hoarseness, fever; and  
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence  
 CC is also used as vaccines for immunising humans against diseases  
 CC caused by C. pneumoniae.  
 XX Sequence 490 AA;  
 SQ  
 Query Match 99.1%; Score 2241; DB 21; Length 490;  
 Best Local Similarity 99.1%; Pred. No. 5.5e-143;  
 Matches 459; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MVNPIGPGPIDETERTPPADLSAAGLEASAAKSAEQAQRIAGAEAKPKESKTSVERWSI 60  
 DB 1 mvnpiqpiddetertppadlsaggleasaanksaegriagaekpketsverwsi 60  
 QY 61 LRSVAVNMLSLADKLGIASSSSSTSRSDVDSTTATAPPPPTSDDYKTQATAYDT 120  
 DB 61 lrsavnlmsladdkligiassssstsrsvdvtatppppptdddyktqataydt 120  
 QY 121 IFTSTSLADIQAALVSLQDAVTNKTAAETAEAEWETKNADAKVGAQITELAKY 180  
 DB 121 iftstsladiqaalvslqdvtnktaataetaeaewetknadavkvgaqitelaky 180  
 QY 181 ASDNQAILDSLGKLTSPDLLQATLQSVANNKKAELLKEMQDNVPVPGKTPAIQAQSLVD 240  
 DB 181 asdnqaildslgkltspdllqatllqsvannnkkaellkemqdnvpvpgktpaiqaqslvd 240  
 QY 241 QTDTATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTQIAEAQ 300  
 DB 241 qtdtatatqiekdgnaigdayfagqnasgavenaksnnsisnidakaaiaataktqiaeq 300  
 QY 361 DDAENETASILMSGFQMIHMFNTENPDSSQAQQLAAQARAAGDSDSAAALADAQK 420  
 DB 361 ddaenetasilmsgfmihmfnntenpdsqaqqllaqaraagdsdaaaladaqk 420  
 QY 421 ALEAALGKAGQGGILNALGQIASNAVVSAGVLPLOQVLIWIRA 463  
 DB 421 aleaalgkagqggilnalgqiasnavvsagvlploqvliwira 463  
 RESULT 2  
 AAY71954  
 ID AAY71954 standard; Protein; 651 AA.  
 XX  
 AC AAY71954;  
 XX  
 DT 26-MAR-2001 (first entry)  
 XX Chlamydia pneumoniae 76 kDa full-length protein.  
 DE  
 XX 76 kDa protein; bactericidal; diagnosis; prevention; treatment;  
 KW

KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;  
 KW acute respiratory disease; cough; sore throat; hoarseness; fever;  
 XX vaccine; immunisation.  
 XX Chlamydia pneumoniae.  
 XX WO2000066739-A2.  
 PN  
 XX 09-NOV-2000.  
 PD  
 XX 03-MAY-2000; 2000WO-CA00511.  
 PF  
 XX 03-MAY-1999; 99US-0132270.  
 PR 30-JUN-1999; 99US-0141276.  
 XX  
 PA (AVET ) AVENTIS PASTEUR LTD.  
 XX  
 XX Murdin AD, Oomen RP, Wang J, Dunn P;  
 PI  
 XX WPI: 2000-687542/67.  
 DR N-PSDB; AAD02063.  
 XX  
 XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,  
 PT useful for vaccinating against Chlamydia infections -  
 PT  
 XX Claim 16a; Fig 1; 112pp; English.  
 PS  
 XX The present sequence is Chlamydia pneumoniae full-length  
 CC 76 kDa protein. C. pneumoniae 76 kDa protein  
 CC is used in the diagnosis, prevention and treatment  
 CC of C. pneumoniae infections (e.g. pneumonia, upper respiratory  
 CC tract disease, bronchitis, sinusitis and acute respiratory  
 CC disease such as cough, sore throat, hoarseness, fever; and  
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence  
 CC is also used as vaccines for immunising humans against diseases  
 CC caused by C. pneumoniae.  
 XX Sequence 651 AA;  
 SQ  
 Query Match 96.8%; Score 2188; DB 21; Length 651;  
 Best Local Similarity 98.9%; Pred. No. 2.9e-139;  
 Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MVNPIGPGPIDETERTPPADLSAAGLEASAAKSAEQAQRIAGAEAKPKESKTSVERWSI 60  
 DB 1 mvnpiqpiddetertppadlsaggleasaanksaegriagaekpketsverwsi 60  
 QY 61 LRSVAVNMLSLADKLGIASSSSSTSRSDVDSTTATAPPPPTSDDYKTQATAYDT 120  
 DB 61 lrsavnlmsladdkligiassssstsrsvdvtatppppptdddyktqataydt 120  
 QY 121 IFTSTSLADIQAALVSLQDAVTNKTAAETAEAEWETKNADAKVGAQITELAKY 180  
 DB 121 iftstsladiqaalvslqdvtnktaataetaeaewetknadavkvgaqitelaky 180  
 QY 181 ASDNQAILDSLGKLTSPDLLQATLQSVANNKKAELLKEMQDNVPVPGKTPAIQAQSLVD 240  
 DB 181 asdnqaildslgkltspdllqatllqsvannnkkaellkemqdnvpvpgktpaiqaqslvd 240  
 QY 241 QTDTATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTQIAEAQ 300  
 DB 241 qtdtatatqiekdgnaigdayfagqnasgavenaksnnsisnidakaaiaataktqiaeq 300  
 QY 301 KKFDPSPILQAEQMIHMFNTENPDSSQAQQLAAQARAAGDSDSAAALADAQK 360  
 DB 301 kkfpdspilqaeqmihmfnntenpdsqaqqllaqaraagdsdaaaladaqk 360  
 QY 361 DDAENETASILMSGFQMIHMFNTENPDSSQAQQLAAQARAAGDSDSAAALADAQK 420  
 DB 361 ddaenetasilmsgfmihmfnntenpdsqaqqllaqaraagdsdaaaladaqk 420  
 QY 421 ALEAALGKAGQGGILNALGQIASNAVVSAGVLP 454



```

Db 421 aleaalgkagggqilnalqgiaseavvsagvpp 454
|||||
RESULT 3
AAY71956
ID AAY71956 standard; Protein: 452 AA.
XX
AC AAY71956;
XX
DT 26-MAR-2001 (first entry)
XX
DE 3'-truncated Chlamydia pneumoniae 76 kDa protein.
XX
KW 76 kDa protein; bactericidal; diagnosis; prevention;
KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
KW acute respiratory disease; cough; sore throat; hoarseness; fever;
KW vaccine; immunisation; treatment; truncation mutant; mucin.
XX
OS Chlamydia pneumoniae.
OS Synthetic.
PN WO200066739-A2.
XX
PD 09-NOV-2000.
XX
PF 03-MAY-2000; 2000WO-CA00511.
XX
PR 03-MAY-1999; 99US-0132270.
PR 30-JUN-1999; 99US-0141276.
XX
PA (AVET ) AVENTIS PASTEUR LTD.
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI: 2000-687542/67.
DR N-FSDB: AAD02065.
XX
PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
PT useful for vaccinating against Chlamydia infections -
XX
PS Claim 16c; Page 104-106; 112pp; English.
XX
CC The present sequence is 3'-truncated
CC Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
CC protein is used in the
CC diagnosis, prevention and treatment of C. pneumoniae
CC infections (e.g. pneumonia, upper respiratory
CC tract disease, bronchitis, sinusitis and acute respiratory
CC disease such as cough, sore throat, hoarseness, fever; and
CC abnormal chest sounds on auscultation). C. pneumoniae sequence
CC is also used as vaccines for immunising humans against diseases
CC caused by C. pneumoniae.
XX
SQ Sequence 452 AA;

Query Match 96.6%; Score 2184; DB 21; Length 452;
Best Local Similarity 99.1%; Pred. No. 3.4e-139;
Matches 448; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MWNPGCPIDETERTPPADLSAQGLESAANKSAEORAGAEAKPKESKTSVERWSI 60
|||||
DB 1 mwnpgpdpidetertppadlsaggleasaanksaegakpkesktidsverwsi 60
|||||
QY 61 LRSVAVNMLSLADKLGIASSSSSTSRSDVDSTATPTPPPTSDYKTQACTAYDT 120
|||||
DB 61 lrsavnlmsladvkigiasssssstsrsvdvtatcptppptfdyktqactaydt 120
|||||
QY 121 IFTSTSLADIQAALVSLQDAVTNKTATDEETAIAAEWETKNADAIVKGAQITELAKY 180
|||||
DB 121 iftstsladiqaalvslqdvatnktatdeetaiaaewetknadvkvgaqitelaky 180
|||||

181 ASDNQAILDSLGKLTSPDLQFALLOQSVANNKKAELLKEMQDNPVVGKTPATAQSLVD 240
|||||
181 asdnqailds1gkltstfollqalqsvannnkaaelikemqdnvvpvgktpataqslvd 240
|||||
241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAIATAKTQIAEAQ 300
|||||
241 qtdatatqiekdgnairdayfagqnasgavenaksnnsisnidsakaaiataktqiaeaq 300
|||||
301 KKFPDSPILQAEQWVIOAEKDLNKNIRPADGSDVPNPGTTVGGSKQOQSSIGSIRVSMLL 360
|||||
301 kkfpdspilqaeqwmviquekdlknikpadgsdvpnpgttvggskqgssigsirvsmll 360
|||||
361 DDAENETASILMSGFROMIHMFNTENPDSSQAQOELAAOARAAKAGDSDSAAALADAOK 420
|||||
361 ddaenetasilmsgfrqmihmfntenpdsgaaqqlaaraqaaagddsaaladaok 420
|||||
421 ALEAALGKAGQOQGIILNALGQIASAAVVSAGV 452
|||||
421 aleaalgkagqggilnalqgiaseavvsagv 452
|||||

RESULT 4
AAY35358
ID AAY35358 standard; Protein: 478 AA.
XX
AC AAY35358;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed polypeptide.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST ) GENSET.
XX
PI Griffais R;
XX
DR WPI: 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1156; Disclosure; 1912pp; English.
XX
CC AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 478 AA;

Query Match 95.6%; Score 2162; DB 20; Length 478;
Best Local Similarity 99.1%; Pred. No. 1.1e-137;

```



Matches 443; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MWNPIGPPIDETERTPPADLSAAGLEASAAKSAEQRITAGAAKPKESKTSVERWSI 60  
 Db 3 mwnpiggpiddetertppadlsaggleasaanksaegrtaagaaekpketsdverwsi 62  
 QY 61 LRSVAVNLSADKLGIASSNSSSTSRSDVDSTTATPTPPPTSDVKYKTAQAYDT 120  
 Db 63 lrsavnlmsladvklgiassnssstsradvdstttatptppptfddyktqagaydt 122  
 QY 121 IFTSLADIQAALVSLQDAYVNIKDTAATDEETAIAAEWETKNADAIVKGAQITELAKY 180  
 Db 123 iftstsladiqaalvslqdayvnikdtaadeetaiaaewetknadvkgaqitelaky 182  
 QY 181 ASDNQAIILDSLGKLTSLFLLQVANNKAAELLKEMQDNPVPGKTPAIAQSLVD 240  
 Db 183 asdnqaildsilgkltslfllqavannkkaellkemqdnvpgktpaiagslvd 242  
 QY 241 OTDATATQTEKDNAGIDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTQIAEAQ 300  
 Db 243 tdatatqtekdngaidayfagqnasgavenaksnnsisnidsakaaiaataktqiaeaq 302  
 QY 301 KKFDPSPILQAEQMVIOAEKDLNKPADGSDVPNPQTTVGGSKQGGSSIGSIRVSMLL 360  
 Db 303 kkfpdspilqaeqmvioaeekdlknkpadgsdvpnpqttvggskqgssigsirvsmll 362  
 QY 361 DDAENETASILMSGFQMIHMFNTENPDSSQAQELAAQAAKAAAGDSDSAAALADAQK 420  
 Db 363 ddaenetasilmsgfmihmfntenpdssqaqelaaqaaakaaagdsdaaaladaqk 422  
 QY 421 ALEAALGKAGQGGILNALGQIASAAV 447  
 Db 423 aleaalgkaggggillnalgqiasaav 449

RESULT 5  
 AAY71955  
 ID AAY71955 standard; Protein; 583 AA.  
 AC AAY71955;  
 XX  
 XX  
 DT 26-MAR-2001 (first entry)  
 DE  
 DE 5'-truncated Chlamydia pneumoniae 76 kDa protein.  
 KW 76 kDa protein; bactericidal; diagnosis; prevention;  
 KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;  
 KW acute respiratory disease; cough; sore throat; hoarseness; fever;  
 KW vaccine; immunisation; treatment; truncation mutant; mutein.  
 XX  
 OS Chlamydia pneumoniae.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 497 /note= "Encoded by ACT"  
 FT  
 XX  
 XX WO2000066739-A2.  
 PN  
 XX  
 PD 09-NOV-2000.  
 XX  
 XX 03-MAY-2000; 2000WO-CA00511.  
 XX  
 XX 03-MAY-1999; 99US-0132270.  
 PR 30-JUN-1999; 99US-0141276.  
 XX  
 XX (AVET ) AVENTIS PASTEUR LTD.  
 PA  
 XX Murdin AD, Oomen RP, Wang J, Dunn P;  
 PI  
 XX WPI: 2000-687542/67.  
 DR N-PSDB; AAD02064.  
 DR  
 XX

PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,  
 PT useful for vaccinating against Chlamydia infections -  
 XX Claim 15b; Page 100-102; 112pp; English.  
 CC The present sequence is 5'-truncated  
 CC Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa  
 CC protein is used in the  
 CC diagnosis, prevention and treatment of C. pneumoniae  
 CC infections (e.g. pneumonia, upper respiratory  
 CC tract disease, bronchitis, sinusitis and acute respiratory  
 CC disease such as cough, sore throat, hoarseness, fever; and  
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence  
 CC is also used as vaccines for immunising humans against diseases  
 CC caused by C. pneumoniae.  
 XX Sequence 583 AA;  
 SQ

Query Match 81.8%; Score 1850; DB 21; Length 583;  
 Best Local Similarity 98.7%; Pred. No. 1.4e-116;  
 Matches 381; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 69 MSLADKLGIASSNSSSTSRSDVDSTTATPTPPPTSDVKYKTAQAYDTIFTSTSLA 128  
 Db 1 msladklgiassnssstsradvdstttatptppptfddyktqagaydtiftstsla 60  
 QY 129 DIQAALVSLQDAYVNIKDTAATDEETAIAAEWETKNADAIVKGAQITELAKYSDNQAIL 188  
 Db 61 diqaalvslqdayvnikdtaadeetaiaaewetknadvkgaqitelakysdnqail 120  
 QY 189 DSLGKLTSLFLLQVANNKAAELLKEMQDNPVPGKTPAIAQSLVDOTDATATQ 248  
 Db 121 dslgkltslfllqavannkkaellkemqdnvpgktpaiagslvdotdatatq 180  
 QY 249 IEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTQIAEAQKFPDPSPI 308  
 Db 181 iekdgnaigdayfagqnasgavenaksnnsisnidsakaaiaataktqiaeaqkfpdsp 240  
 QY 309 LOEAQMVIOAEKDLNKPADGSDVPNPQTTVGGSKQGGSSIGSIRVSMLLDAENETA 368  
 Db 241 lqeaqmvioaeekdlknkpadgsdvpnpqttvggskqgssigsirvsmlldaeneta 300  
 QY 369 SILMSGFQMIHMFNTENPDSSQAQELAAQAAKAAAGDSDSAAALADAQKALEAALGK 428  
 Db 301 silmsgfmihmfntenpdssqaqelaaqaaakaaagdsdaaaladaqkaleaalgk 360  
 QY 429 AGQOQGIILNALGQIASAAVSGVLP 454  
 Db 361 agqqggilnalgqiasaavsvagvlp 386

RESULT 6  
 AAY37571  
 ID AAY37571 standard; Protein; 350 AA.  
 XX  
 AC AAY37571;  
 XX  
 XX 07-OCT-1999 (first entry)  
 DT  
 XX Chlamydia trachomatis surface exposed protein.  
 XX  
 XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 XX WO9928475-A2.  
 XX  
 XX 10-JUN-1999.  
 PD  
 XX



DE Chlamydia sp. protein # 6.

DT

05-SEP-2001 (first entry)



```

XX DE Protein encoded by Chlamydia trachomatis clone CT622.
XX KW Chlamydia; vaccine; infection; fusion protein; antigen;
XX KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
XX KW acute respiratory tract infection; CapI; CT529; OMCB;
XX KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.
XX OS Chlamydia trachomatis.
XX PN WO200140474-A2.
XX PD 07-JUN-2001.
XX PF 04-DEC-2000; 2000WO-US32919.
XX PR 03-DEC-1999; 99US-0454684.
XX PR 19-APR-2000; 2000US-0556877.
XX PR 20-JUN-2000; 2000US-0598419.
XX PA (CORI-) CORIXA CORP.
XX PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;
XX DR WPI; 2001-374831/39.
XX CH Chlamydia polypeptides and fusion proteins useful for preventing pelvic
XX PT inflammatory disease, trachoma, acute respiratory tract infections,
XX PT atherosclerosis and heart disease -
XX PS Claim 66; Page 257-259; 295pp; English.
XX CC The present sequence is provided in a specification relating to
XX CC compounds and methods for the treatment and diagnosis of chlamydial
XX CC infection. The compounds provided include polypeptides and fusion
XX CC proteins comprising immunogenic portions of Chlamydia antigens
XX CC and DNA sequences encoding such polypeptides. They are useful for
XX CC vaccinating against chlamydial infection, which causes pelvic
XX CC inflammatory disease, trachoma, acute respiratory tract infections,
XX CC atherosclerosis and heart disease.
XX SQ Sequence 361 AA;

Query Match 10.2%; Score 231.5; DB 22; Length 361;
Best Local Similarity 35.4%; Pred. No. 5.8e-08;
Matches 57; Conservative 30; Mismatches 61; Indels 13; Gaps 5;

QY 301 KKEPSPILQEAQVIOAEKDLNKPAGSDVNPQTTGGSKQGGSSIGSI----- 354
Db 2 qeIadspIvkkaeeqinqdqdtitps-glIdipIvqps-gsaasagsaagalkssnns 59
QY 355 -RVSMILDDAENPTASILMSGPROMIHTENTPDSQAQQLAAQARAARAK---AAGDSD 410
Db 60 grIsIlldvdvemaIamqgfIrmleqfInvnnpatakelqameqltmsdqIvgadge 119
QY 411 AAALADAGKALEAALGKAGQOQQQILNALGQIASAAVVSAG 451
Db 120 lpaeqIaIkdaIaqaI-kpsadglatamgqvafaakvvg 159

RESULT 9
AAR27745
ID AAR27745 standard; Protein; 1822 AA.
XX AC AAR27745;
XX XX
XX DT 04-MAR-1993 (first entry)
XX DE Extracellular factor related protein.
XX KW EF*; detection; prevention; screening; diagnostic.
XX XX

```

```

OS Streptococcus suis type II (non-pathogenic).
XX FH Key Location/Qualifiers
XX FT Peptide 1..46
XX FT Peptide /note= "signal peptide"
XX FT Peptide 47..1822
XX FT Region /note= "mature peptide"
XX FT Region 858..861
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 934..937
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 999..1002
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 1075..1078
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 1264..1267
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 1362..1365
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 1438..1441
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 1514..1517
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 1590..1593
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 1666..1669
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX FT 1741..1744
XX FT /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX PN WO9216630-A.
XX XX
XX PD 01-OCT-1992.
XX PF 19-MAR-1992; 92WO-NL000054.
XX PR 21-MAR-1991; 91NL-0000510.
XX PA (DIER-) CENT DIERGENEESKUNDIG INST.
XX PI Smith HE, Vecht U;
XX DR WPI; 1992-349215/42.
XX DR N-PSDB; AAQ29471.
XX CC Deoxyribonucleic acid encoding virulence characteristic of
XX CC Streptococcus suis - useful for antibody and polypeptide for
XX CC diagnosing and preventing infections in pigs and humans
XX PS Claim 9; Fig 1b; 86pp; English.
XX CC The sequence is that of the extracellular factor related protein
XX CC from Streptococcus suis type II (non-pathogenic) which allows the
XX CC detection and the prevention of infections by S. suis in a more
XX CC effective manner than was previously possible. It facilitates
XX CC screening of e.g. pigs and elimination of infected and carrier pigs
XX CC can then be carried out. The new diagnostic tests can distinguish
XX CC between avirulent and virulent strains. It may be used in the prodn.
XX CC of a vaccine. See also AAR27744 and AAR27746.
XX SQ Sequence 1822 AA;

Query Match 8.0%; Score 181.5; DB 13; Length 1822;
Best Local Similarity 23.9%; Pred. No. 0.0011;
Matches 117; Conservative 78; Mismatches 175; Indels 119; Gaps 27;

QY 10 IDETERTPPADLSAAGLEASAAKSAEQAQRIAGAEAKPKESKTSV---ERWSILESAAVN 66
Db 1072 IdDnpnltpde-----kesaknaveeaaakvata-aidkastpdavqveedkgv--aaIn 1122
QY 67 ALMSIADKLGIATSSNSSSTSRSD-----VDSTTATATPPPTSDDYKTAQTAY 118

```



Db 1123 litakadagviaakladielkkedkqaeakaidastmt-----neekaiakkaal 1173  
Qy 119 DTFTSLADIAQALVSLQDA---VTN- IKDTAATDEETAIABEWETKNADAIKVGAGI 174  
Db 1174 qdv-----vdkgkaeadaarvatneiheattekakaealageksltdtkearda 1225  
Qy 175 TELAKYASDQAILDSLGK--LTSFDLLQTLALLOSVAANNKAAELLLKEMQDNVPVPGK-- 230  
Db 1226 velak---dke-----lgkeairteeaatkiveklaedtr-----kaiednplnsdedk 1272  
Qy 231 -----TPATAQSLV---DOTDATATOIEKDGNAIGDAYFAGQNASGAVENAKSNNSIS 280  
Db 1273 qaeikkltdavaktiatmrdnadrtdkrtqaeak-aqalad-----lekaketqkia 1320  
Qy 281 NIDSAKAI-----AT---AKTOIAE-----AOKFPDPSPILOEAOQVIOAE 320  
Db 1321 d-----kaaidrltlivkgeleatkqdktkiakdaaaakeaisnplntdaekktfda 1376  
Qy 321 KDLKNIKPADGSDVPNPOTTTVGGSKQGGSSIGSIRVSM-----LDAENETASILMSGPRQ 377  
Db 1377 vdaevakandaiksaatspadv--qkeedagvaiaaedvldaakqdknkiakdaaaa-ke 1433  
Qy 378 MIHMENTENPDQSQAQOEL---AQARAKAAGDDSAALADADAQKALEAALGKAGQOQG 434  
Db 1434 ai-----gsnplntdaekktftdavadavakandaiksaatspadvqkeeda--gvaiaaed 1487  
Qy 435 ILNALGQIA 443  
Db 1488 vldaakqda 1496

## RESULT 10

AAV37572  
ID AAV37572 standard; Protein; 331 AA.

XX  
AC AAV37572;

XX  
DT 07-OCT-1999 (first entry)

XX  
DE Chlamydia trachomatis cellular envelope protein.

XX  
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX  
OS Chlamydia trachomatis.

XX  
PN WO9928475-A2.

XX  
PD 10-JUN-1999.

XX  
PF 27-NOV-1998; 98WO-IB01939.

XX  
PR 04-NOV-1998; 98US-0107077.

XX  
PR 28-NOV-1997; 97FR-0015041.

XX  
PR 17-DEC-1997; 97FR-0016034.

XX  
PA (GEST ) GENSET.

XX  
PI Griffais R;

XX  
DR WPI; 1999-371125/31.

XX  
PT Genome sequence of Chlamydia trachomatis

XX  
PS Disclosure; Page 1227-1228; 1755pp; English.

XX  
CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
CC can also be used to control growth of the microorganism. Chlamydia  
CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,  
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
CC perihepatitis, bartholinitis; pneumopathy in breast feeding infants;  
CC and venereal lymphogranulomatosis. The polypeptides of the invention  
CC may be of use in treating these diseases.

XX  
SQ Sequence 331 AA;

Query Match 7.3%; Score 164.5; DB 20; Length 331;  
Best Local Similarity 38.5%; Pred. No. 0.0017;  
Matches 45; Conservative 12; Mismatches 53; Indels 7; Gaps 3;

Qy 338 GTTVGGSKQGGSSIGSIRVSMILLDDAENETASILMSGFROMIHMENTENPDQSQAQOELA 397

Db 17 gsaagalk---ssnsggrisllldvdnemaaiamqgfismieqfnnvmpatakelqame 73

Qy 398 AQARAAK---AAGDDSAALADAKALEAALGKAGQOQGGILNALGQIASAAVVSAG 451

Db 74 aqltamsqqlvgadgelpaeiqaikdalaqal-kpsadglatamgqvafaaakvvg 129

## RESULT 11

AAV81609

ID AAV81609 standard; Protein; 1237 AA.

XX  
AC AAV81609;

XX  
DT 24-MAY-2000 (first entry)

XX  
DE Streptococcus pneumoniae type 4 protein sequence #109.

XX  
KW Streptococcus pneumoniae; vaccine; screening; protein antigen;  
KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
KW pneumococcal disease.

XX  
OS Streptococcus pneumoniae.

XX  
PN WO200005737-A2.

XX  
PD 10-FEB-2000.

XX  
PF 27-JUL-1999; 99WO-GB02451.

XX  
PR 27-JUL-1998; 98GB-0016337.

XX  
PR 19-MAR-1999; 99US-0125164.

XX  
PA (MTCR-) MICROBIAL TECHNIQS LTD.

XX  
PI Gilbert CFG, Hansbro PM;

XX  
DR WPI; 2000-195300/17.

XX  
PT New Streptococcal protein, useful as a vaccine, for diagnosis of  
PT pneumococcal diseases and for screening agents capable of antagonizing  
PT or inhibiting expression of the protein

XX  
PS Claim 1: Page 83; 108pp; English.

XX  
CC AAY81501 to AAY81679 represent specifically claimed protein sequences  
CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent  
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.  
CC The sequences have antibacterial and antiinflammatory properties.  
CC The protein sequences, and fragments of them, are useful as immunogens  
CC and/or antigens. The nucleotide sequences can be used in vaccines and in  
CC diagnostic assays. The proteins and nucleotides can be useful for the  
CC detection and diagnosis of S. pneumoniae. The protein sequences are also  
CC useful for screening an agent capable of antagonizing, inhibiting or  
CC interfering with the function or expression of the proteins in which the  
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection  
CC and meningitis. AAA05591 to AAA05614 represent primers used in the  
CC exemplification of the present invention.



```

XX 50 Sequence 1237 AA;
      Query Match 6.8%; Score 153.5; DB 21; Length 1237;
      Best Local Similarity 19.9%; Pred. No. 0.051;
      Matches 84; Conservative 84; Mismatches 205; Indels 49; Gaps 9;

Qy 13 TERTPPADLSAQGLEASANKSAAQRIAGAEKPKESKTDTSVERWILSAVNALMSLA 72
Db 467 tsasasasasasasasasasasasasasasasasasasasasasasasasas 526
Qy 73 DKLGTAGSSSSSTSRGADVDSITATAPPTPPPTSDDYKTAQATYDTITFTSLADIQA 132
Db 527 ---gsaststasasasasasasasasasasasasasasasasasasasasas 577
Qy 133 ALVSLQDAVNIKTAATDETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDSL 192
Db 578 stasasasasasasasasasasasasasasasasasasasasasasasasas 634
Qy 193 KLTSPDLLQTLALQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVDQTDATATQIEKD 252
Db 635 asasasasasasasasasasasasasasasasasasasasasasasasasas 686
Qy 253 GNAIGDAYFAGQNASGAVENAKSNNSISNIDSAK-AAATATAKTAQIAEAKKFPDPSILQE 311
Db 687 astasasasasasasasasasasasasasasasasasasasasasasasasas 741
Qy 312 AEQWVIAEKDLNKKIPADGSDVPNPGTGVGSKQGGSSI-GSTRVSMLLDDAENETASI 370
Db 742 -----sastsasasasasasasasasasasasasasasasasasasasas 785
Qy 371 LMSGFRQMIHMFNTENPDPSQAQOELAAQARAAGDSDSAAALADAQAALCKAG 430
Db 786 sas-----esastsasasasasasasasasasasasasasasasasasas 839
Qy 431 QQ 432
Db 840 rq 841

RESULT 12
AA991443
ID AA991443 standard; Protein; 600 AA.
XX AC AA991443;
XX 26-SEP-2001 (first entry)
XX C glutamic protein fragment SEQ ID NO: 5197.
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX Corynebacterium glutamicum.
XX EP1108790-A2.
XX 20-JUN-2001.
XX 18-DEC-2000; 2000EP-0127688.
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
XX N-PSDB; AAH66662.

```

```

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 5197; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX Sequence 600 AA;
XX
      Query Match 6.8%; Score 153; DB 22; Length 600;
      Best Local Similarity 21.6%; Pred. No. 0.022;
      Matches 96; Conservative 66; Mismatches 187; Indels 96; Gaps 14;

Qy 18 PADLSAQGLEASANKSAAQRIAGAEKPKESKTSVE-RWILSAVN-ALMSLADKL 75
Db 24 pddaaiaqaeevnsagdgevarlags-lsstdaeinrvelmgalreevnxslvldhaq 82
Qy 76 GIASSNSSSTSRGADVDSITATAPPTPPPTSDDYKTAQATYDTITFTSLADIQAALV 135
Db 83 aiae-----qarqdaaakkdlddsaqie 107
Qy 136 SLQDAVTNIKDTAATDETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDSLGLT 195
Db 108 aagerldeisraayrqnqgtskigsgnngusedaldrqtylrtsaekqqaaveeldrlr 167
Qy 196 SFDLLQTLALQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVDQTDATATQIEKDGA 255
Db 168 t-----enankesvlrq-----arivaegreaavekvqvtca 200
Qy 256 IGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTAQIAEAKKFPDPSILQEAEQM 315
Db 201 ---aianseqlnvltnrs-tlvaqrdgaernlalaaraqadnlqgqraeyeeffqaeqa 256
Qy 316 VIOAEKD----LKNIKPADGSDVPNPGTGVGSKQGGSSIGSTRVSMLLDDAENETASIL 371
Db 257 riqaeaeaaqaeeekrradea-aaqaaaeaaqaaqaaeeaaqaaqaaqaaqaaqaaqaa 314
Qy 372 MSGFRQMIHMFNTENPDPSQAQOELAAQARAAGDSDSAAALADAQAALCKALEAA 425
Db 315 -----aetqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqa 362
Qy 426 LGKAGQQOQILNALGQIASAAVUSA 450
Db 363 -aqalreqaltaa--siaaaaliaa 384

RESULT 13
AAW02098
ID AAW02098 standard; Protein; 1561 AA.
XX AC AAW02098;
XX 27-OCT-1996 (first entry)
XX S. mutans antigen I/II.
XX Caries; antigen I/II; epitope; tooth decay; vaccine; immunisation.

```



OS Streptococcus mutans.  
 XX Key Location/Qualifiers  
 FH Misc-difference 618..650  
 FT /note= "amino acids 618-650 differ from the  
 FT residues deduced from the nucleic acid  
 FT sequence."  
 XX W09623886-A1.  
 XX 08-AUG-1996.  
 PD 31-JAN-1996; 96WO-CB00207.  
 XX 31-JAN-1995; 95GB-0001826.  
 PR (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 PA Kelly C, Lehner T;  
 XX WPI: 1996-371434/37.  
 XX N-PSDB; AAT36122.  
 DR Nucleic acid encoding polypeptide for prevention or treatment of  
 XX dental caries - which stimulates T or B cell response, and/or  
 PT adheres to tooth in competition with Streptococcus mutans antigen  
 PT I/II  
 XX Disclosure; Page 45-46; 63pp; English.  
 PS Streptococcus mutans antigen I/II (AAW02098) is a 185 kDa cell surface  
 CC protein at least partly responsible for S. mutans adhesion to teeth.  
 CC The I/II antigen includes a series of overlapping T-cell, B-cell and  
 CC adhesion epitopes. Fragments (see also AAT36111-21) of the I/II  
 CC antigen gene (see also AAT36122) can be used to produce recombinant  
 CC polypeptides (AAW02087-97) carrying such epitopes for use in vaccines  
 CC for immunisation against dental caries.  
 XX Sequence 1561 AA;

Query Match 6.7%; Score 152; DB 17; Length 1561;  
 Best Local Similarity 22.3%; Pred. No. 0.087;  
 Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;  
 QY 48 KESKTSVERWSILRSANVNAALMSLADKLGIASSN-SSSSTSRSDVDSTATATPTPPPT 106  
 DB 2 kvkkygfrkskistlcvagltvvaavsvagkvfadetttsdvdkvvgqtgnpat 61  
 QY 107 -----SDDYKTOAQATYDIFTST-----SLADTQAAALVS-LQDAVNTIKDT 147  
 DB 62 nlpeaagsaskqaedsqtklerqmvhtievptktdldqaaakdaksagvnnvgdadvn-kg 120  
 QY 148 AATDEETATAAEWE-----TKNADAIK-----VGAQITELAKYASDNOAILDSLGK 193  
 DB 121 vtace-avqketelekedytkqaedkktdtdykdsvaheaeavakikakngatkeqyqk 179  
 QY 194 LFSFOLL-OTALLQSVANNKKAEL-----LKEMQDNVPVPOKTPAQSLSVDQTDATAT 247  
 DB 180 ----dmvnhkaeverinaanaasakayeklaqyqadlaavqktnaanqasy-qkalaay 234  
 QY 248 QIE----KQNAIGDAYFAGQNASGAVENAKSNNSISNIDSASAKAIATAK-----TQIA 299  
 DB 235 qaelkrvgeanaakaay---dtavaannakntelaaneelkrnatakaeyeklaqy 291  
 QY 300 Q---KKFPDPIQAE-OMVIOA-EKDLKNIKPAD-----GSDVPNPGTTVG 342  
 DB 292 qaelkrvgeanaaneadyqakltayqtelarvqkanadakaayaevaannaknaaltae 351  
 QY 343 GS-----KQGSISGIRVSMILLDAENEFASILMSGFROMIHFNTE 385  
 DB 352 ntaikqrnenakatyeealkyeadlaavkvkanaaneadyqak---ltayqtelarvqka 408

QY 386 NFDSSAAQOELAAQAARAKAA--GDSAAAL-ADAQKALEALCK 428  
 DB 409 nadakaayaevaannaaltaentaikkrnadakadyeaklak 454  
 RESULT 14  
 AAR84568  
 ID AAR84568 standard; Protein; 643 AA.  
 XX AAR84568;  
 AC 09-MAY-1996 (first entry)  
 DT Trypanosoma cruzi TCR27 polypeptide, Ag15.  
 DE TCR27; Chagas disease; repeat unit; diagnosis; blood screening;  
 XX recombinant; fusion protein; glutathione-S-transferase.  
 KW Trypanosoma cruzi.  
 OS Trypanosoma cruzi.  
 XX Key Location/Qualifiers  
 FH 329..552  
 FT /label= repeat\_region  
 FT /note= "16 of 69 repeat units of 14 amino acids"  
 XX W09525797-A1.  
 PN 28-SEP-1995.  
 PD 20-MAR-1995; 95WO-US03191.  
 XX 24-MAR-1994; 94US-0216894.  
 PA (KIRC/) KIRCHHOFF L V.  
 PA (ORSU/) OFSU K.  
 XX Kirchoff LV, Otsu K;  
 XX WPI: 1995-344618/44.  
 DR N-PSDB; AAT05332.  
 XX New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi - as  
 PT immunoassay reagent for specific diagnosis of Chagas disease, also  
 PT related nucleic acid and transformed cells  
 XX Disclosure; Page 40-41; 68pp; English.  
 CC AAR84565-R84569 are polypeptides of the TCR27 protein of T. cruzi  
 CC The proteins are all fusion products with glutathione-S-transferase  
 CC (GST) and some contain a linker sequence. The TCR27 protein comprises  
 CC a 95 amino acid (aa) N-terminal region; 69 repeats of a highly  
 CC conserved 14 aa sequence and a 68 aa C-terminal region. This sequence  
 CC encodes the GST sequence, the A944 polypeptide contg. 16 of the 69  
 CC repeat units and also contains the amino and carboxy terminal  
 CC peptides of TCR27. The TCR27 polypeptides of the invention are useful  
 CC for the diagnosis of Chagas disease (American Trypanosomiasis), they  
 CC are capable of detecting anti-T. cruzi antibodies; or for blood  
 CC screening. The TCR27 protein has epitopes to which most T. cruzi  
 CC infected individuals have antibodies. The TCR27 polypeptides will not  
 CC react with serum from patients with leishmaniasis; schistosomiasis,  
 CC or autoimmune disease and are hence less likely to cause false  
 CC positives in diagnosis.  
 XX Sequence 643 AA;

Query Match 6.7%; Score 150.5; DB 16; Length 643;  
 Best Local Similarity 21.5%; Pred. No. 0.035;  
 Matches 99; Conservative 58; Mismatches 169; Indels 135; Gaps 17;  
 QY 7 PGPIDETPTPADISAQGLEASAAKSAQAQRIAGAEAKPKESKTSVERWSILRSANV 66



Db 216 ppskdlvprgsqll--qgaennitnskemtikrekvvkkaekelkdainratkleeern 273  
 QY 67 ALMSLADKLGIASSNSSSTSRSDVDSTTATPTPPPTSDDYKTOAQATYDTFTSTS 126  
 Db 274 qaykaahk---aeekaktfgrliltfesenlnlkrp----- 307  
 QY 127 LADIAQALVSLQDAVNTN---IKDTAATDE-----ETAIAAEWETKNADAIVGQAQI 174  
 Db 308 -----ndavsnrdkkknsetaktdevekraaeaaakavetekqra-----aea 350  
 QY 175 TELAKYASDNQAILDSLGLKTSFDLLQTALLQSVANNNAKAEELLKEMQDNVVPVPGKTPAI 234  
 Db 351 tkvae-----akrkkaaeaaakavetekqraaeatk 381  
 QY 235 A-----QSLVDQTDATATQIEKDNIGNAIGDAYFAGQNASCAVENAKSN-----NSISNID 283  
 Db 382 aeekqkaaeaaakavetekqraaeatkvaeeakqraaeamkvaeeakqkaaeatkvaee 441  
 QY 284 SAKAAATATAKTQIAEAQR-KFPDSPILQEAQOM-----VIOAEKDLKNIKPADGSDVP 335  
 Db 442 kqkaaea---tkvaeaekqkaaeatkvaeeakqkaaeatkvaeeakq-----kaaeatkva 494  
 QY 336 NPGTTVGSGKQSGSIGIRVSMLLDDAENETASTILMSGFROMIHMFNTENPDSQAAQOE 395  
 Db 495 e-----aekqkaaeatkva-----eaekqkaaeatk-----vaeaekqkaaeatk 533  
 QY 396 LAAQARAARAKAGDDSAALADAKALEALGKAGQOQOIL 436  
 Db 534 kvaeaekqkaa---eatkvaeeakqkaaeakamesqkqrl 572

RESULT 15  
 AAB19849  
 ID AAB19849 standard; Protein; 729 AA.  
 XX  
 AC AAB19849;  
 XX  
 DT 05-MAR-2001 (first entry)  
 XX  
 DE Mycobacterium tuberculosis protein MTB8.  
 XX  
 KW MTB8; tuberculosis; BCG; vaccine; infection; diagnosis.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO2000066157-A1.  
 XX  
 PD 09-NOV-2000.  
 XX  
 PF 04-MAY-2000; 2000WO-US12257.  
 XX  
 PR 04-MAY-1999; 99US-0132505.  
 XX  
 PA (PUBL-) PUBLIC HEALTH RES INST NEW YORK.  
 XX  
 PI Gennaro ML;  
 XX  
 DR WPI; 2001-007153/01.  
 DR N-PSDB; AAA89042.  
 XX  
 PT Novel polypeptide encoded by open reading frames present in  
 PT Mycobacterium tuberculosis genome and not by the BCG strain of M.  
 PT bovis, useful as vaccine and for diagnosing tuberculosis infection  
 XX  
 PS Claim 11; Fig 1; 35pp; English.  
 XX  
 CC The present sequence is that of the Mycobacterium tuberculosis  
 CC MTB8 protein. This is 1 of 8 proteins, i.e. MTB8-1 (see  
 CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)  
 CC identified as being present in the genome of M. tuberculosis but  
 CC absent from the genome of the BCG strain of Mycobacterium bovis.  
 CC MTB8-1 represent reagents that are useful in discriminating between  
 CC M. tuberculosis and BCG and, in particular, for diagnostic methods

CC which discriminate between exposure of a subject to M. tuberculosis  
 CC and vaccination with BCG. The invention features these MTB8  
 CC polypeptides, functional fragments of them, DNA encoding them,  
 CC vectors, transformed cells, and diagnostic, therapeutic, and  
 CC prophylactic (vaccine) methods, including genetic vaccination  
 CC methods.  
 XX  
 SQ Sequence 729 AA;  
 1  
 Query Match 6.5%; Score 148; DB 22; Length 729;  
 Best Local Similarity 23.2%; Pred. No. 0.06;  
 Matches 103; Conservative 46; Mismatches 181; Indels 114; Gaps 17;  
 QY 6 GPGPIDETER--TPADLSAQGLEASAAANKSAEQAQRTAGAEAKPKESKTSVERWSILRS 63  
 Db 351 gpahadesaasvtppaasgvpgraaapsgta---vgaga-----rs 391  
 QY 64 AVNALMSLADKLGIASSNSSSTSRSDVDSTTA---TAPTPPPPTSD--DYKTOAQTA 117  
 Db 392 svgtaaasgagshaatgrapvatsdkaapstraasartapparpsthdhdkpdrsesa 451  
 QY 118 YDTFTSTSLADIQAALVSLQDAVNTNKTDAATDEETAIAAEWETKNADAIVGQAQITEL 177  
 Db 452 ddg--tpvsmipvsaa-----raardaataaasargrgrgdalrria-a 495  
 QY 178 AKVASDNOAILDSLGLKTSFDLLQTALLQSVANNNAKAEELLKEMQDNVVPVPGKTPAIQAS 237  
 Db 496 alnasdna-----gdygffwi----- 512  
 QY 238 LVDQTDATATQIEKDG-----NAIGDAYFA-GQNASGAVENAKSNNSISNIDSAKAIAT 291  
 Db 513 -----tavtttdgsivvansyglavipdgmelpnkvyilasadhaip-vdei-arcat 561  
 QY 292 AKTOIAEAQKKFPD---SPILQEAQOMVIOAEKDLKNIKPADGSDVPNPGTTVGSKQO- 347  
 Db 562 ypvlavqawaafhdmtlravigtaeqlassdpgvakivlepd--dipesgkmtgrslev 619  
 QY 348 -----GSSIGSIRVSMLLDDAENETASTILMSGFROMIHMFNTENPDSQAAQOELAAQARA 402  
 Db 620 vdpasaaqladttdqrlldllppavdvnpvgderhmlwfeilmkpmststatgreaahlra 679  
 QY 403 AKAAAGDDSAALADAKALEAAL 426  
 Db 680 frayaahsqeialhqahtatdaav 703

Search completed: February 7, 2002, 21:34:38  
 Job time: 20757 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:36:19 ; Search time 85.91 Seconds

(without alignments)  
121,278 Million cell updates/sec

Title: US-09-391-606-9

Perfect score: 2261

Sequence: 1 MWNPIGPGIDETERRPPAD.....SAAYVSAGVPLQCVLMIRA 463

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	152	6.7	1561	3	US-08-894-017-23
2	150.5	6.7	643	2	US-08-216-894-8
3	150.5	6.7	643	4	US-09-115-746-8
4	146.5	6.5	1566	2	US-08-687-956A-23
5	143	6.3	564	2	US-08-216-894-2
6	138.5	6.1	564	4	US-07-741-940-2
7	138.5	6.1	564	1	US-08-452-654-2
8	138	6.1	2842	1	US-08-289-548A-7
9	138	6.1	2842	1	US-08-452-654-7
10	138	6.1	2842	1	US-07-741-940-2
11	138	6.1	2843	1	US-08-289-548A-2
12	138	6.1	2843	1	US-08-452-654-2
13	138	6.1	2843	1	US-08-452-655B-2
14	138	6.1	2843	1	US-08-452-655B-7
15	138	6.1	2843	1	US-08-370-235A-2
16	138	6.1	2843	3	US-08-450-582-2
17	138	6.1	2843	3	US-08-450-582-7
18	138	6.1	2973	2	US-08-821-355A-7
19	138	6.1	2973	2	US-09-003-687A-7
20	138	6.1	2973	4	US-09-136-605-7
21	136.5	6.0	2285	4	US-09-308-375-2
22	136.5	6.0	639	1	US-08-466-390-2
23	135.5	6.0	639	1	US-08-470-950-2
24	135.5	6.0	639	1	US-08-467-781-2
25	135.5	6.0	639	1	US-08-195-487-2
26	135.5	6.0	639	2	US-08-483-924-2
27	135.5	6.0	639	2	US-08-483-924-2

28	135.5	6.0	639	5	PCT-US93-06160-2	Sequence 2, Appl
29	131.5	5.8	593	2	US-08-591-079-8	Sequence 8, Appl
30	131.5	5.8	593	2	US-08-591-079-10	Sequence 10, Appl
31	127.5	5.6	635	4	US-08-931-608A-5	Sequence 5, Appl
32	126	5.6	466	4	US-08-235-836C-107	Sequence 107, App
33	123.5	5.5	461	1	US-08-186-222-2	Sequence 2, Appl
34	123.5	5.5	619	2	US-08-372-652-5	Sequence 5, Appl
35	123.5	5.5	619	5	PCT-US95-16311-5	Sequence 2, Appl
36	123.5	5.5	933	3	US-08-293-728-2	Sequence 5, Appl
37	123.5	5.5	933	4	US-09-421-868-2	Sequence 2, Appl
38	123.5	5.5	2756	1	US-08-752-929-11	Sequence 11, Appl
39	123.5	5.5	2756	1	US-08-752-929-11	Sequence 11, Appl
40	123.5	5.5	2756	4	US-09-090-793-7	Sequence 7, Appl
41	123	5.4	1912	1	US-08-409-995-4	Sequence 4, Appl
42	123	5.4	1912	3	US-08-685-467-4	Sequence 4, Appl
43	123	5.4	2353	4	US-09-377-155-33	Sequence 33, Appl
44	123	5.4	2353	4	US-08-913-942-4	Sequence 4, Appl
45	122.5	5.4	1196	1	US-08-144-121-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-08-894-017-23  
Sequence 23, Application US/08894017  
Patent No. 6024958  
GENERAL INFORMATION:  
APPLICANT: Lehner, Thomas  
TITLE OF INVENTION: POLYPEPTIDE FRAGMENTS CAPABLE  
OF COMPETITION WITH STREPTOCOCCUS MUTANS ANTIGEN I/II  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Avenue, NW, suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,017  
FILING DATE: 31-JUL-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB96/00207  
FILING DATE: 31-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Murashige, Kate H  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 25150-20067.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-887-1500  
TELEFAX: 202-822-0168  
TELEX:  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1561 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-894-017-23  
Query Match 6.7%; Score 152; DB 3; Length 1561;  
Best local Similarity 22.3%; Pred. No. 0.004;  
Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;



```

OY 48 KESTDSVERMSILRSVAVNALMSLADKGIASSN--SSSSTSRSDVDSTTATPTPPPT 106
DB 2 KKKTKYGERKSKISTKLCAGVTAASVAVGOKVFADETTTSDVTRKVGSTGTGNPAT 61
OY 107 ---SDDKYTOQATVADTIFST-----SLADIQALVS-LDPAVTNIKDT 147
DB 62 NLEPAQSGSKAQEOSQTKLERQWHTIEVPTDLQAKAKSAGVNVQDADVN-KGT 120
OY 148 AATDEETAIAMEW-----TKNADAIK-----VGAQITELAKVASDQALIDSLGK 193
DB 121 VTAEE-AVQKETEIKEDTQKQEDIKTKTDQKSDVAHAEVAEKIKAKNOATKEQYK 179
OY 194 LTFPLL-QTALLQSVANNKKAEL-----LKEMQDNVPEKTKPAILQSLVDQDATAT 247
DB 180 ---DMVAHKAVERINNAANAASKTAYEAKLAQYQADLAAYOKTUNAAQASY-QKALAAY 234
OY 248 QIE---KQGNALGDAYFAGNAGAVENAKSNNSISNIDSAKAIATATK---TOIAEA 299
DB 235 QAEKLRVQENANAAKAAAY---DTAVAANNAKTEITAAANEIRKNATKAEYETKLAQY 291
OY 300 Q---KKPDSPILOEAE-OMVIOA-EKDKINIKPAD-----GSDVNPQGTTVG 342
DB 292 QAEKLRVQENANANADYQAKLTAYQTELARVQKANADAKAAVEAANAANNKNAALTAE 351
OY 343 GS-----KQOGSSIGSTRVSMLLDDAENETASTILMSFQIMHMENTE 385
DB 352 NTAIQRNENAKATYEALKEADIAAVKKNANAEADYQAK---LTAAYQTELARVOKA 408
OY 386 NPDQAOQOELAAQARAQAA--GDDSAALAL-ADAOKALEALG 428
DB 409 NADAKAAVEAAYANNAANAALTAENTATKKNADAKADYEAKLAK 454

```

# RESULT 2

US-08-216-894-8  
Sequence 8, Application US/08216894  
Patent No. 5876734

## GENERAL INFORMATION:

APPLICANT: Kirchhoff, Louis V.  
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/216,894  
FILING DATE: 24-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-216-894-8

Query Match 6.7%; Score 150.5; DB 2; Length 643;  
Best Local Similarity 21.5%; Pred. No. 0.0015;  
Matches 99; Conservative 58; Mismatches 169; Indels 135; Gaps 17;

```

OY 7 PGPIDETERTPADISAGGLEASANKSAEORLAGAEKESKTSVERMSILRSVAVN 66
DB 216 PPKSDLPVPGSPQUL--QQAENNTNNSKKEKTKLEKYKKAKEKLDAINATKLEERN 273
OY 67 ALMSLADKLGIASSNSSSSTSRSDVDSTTATPTPPPTSDYKQATVADTIFSTIS 126
DB 274 QAYKAHAK--AEKKAATFQRLTFESENINLKKRP----- 307
OY 127 LADIOALVLSQDAVTN-----IKDTAATDE---ETIAAEWETKNADAIVKGAQI 174
DB 308 ---NDAYSNDRKKNSSETAKTDEVEKORAAEAARAAVETEKORA-----AEA 350
OY 175 TELAKVASDQALIDSLGKLTSPDLLQTALLQSVANNKKAELLKEMQDNVVPVGPKPPI 234
DB 351 TKVAE-----AEKRAADAAKAAVETEKORAAEATKV 381
OY 235 A---QSLVDQDATATQIEKQGNALGDAYFAGNAGAVENAKSN-----NSISNID 283
DB 382 AEAEKQKAAEAQAKAVETEKORAAEATKVAAEAEKQAAALAMVAEAEKQAAEATKVAAE 441
OY 284 SAKAIIATATQIAEPAQ-KPPDSPILOEADOM-----VIOAEKDKINIKPADGSDVP 335
DB 442 KQKAAEA--TKVAEAEKQKAAEAATKVAAEAEKQAAEATKVAAEAEKQ---KAAEATKV 494
OY 336 NGTGVGSGKQOGSSIGSTRVSMLLDDAENETASTILMSFQIMHMENTENPDQSOAQOE 395
DB 495 E-----AEKQKAAEATVVA-----EAEKQKAAEATV-----VAEAEKQKAAEAT 533
OY 396 LAAQARAQAAAGDSSAAALADAOKALEALGKAGQOQGITL 436
DB 534 KVAEAEKQKAA--EATKVAAEAEKQKAAEAARAAKAMESQKORFL 572

```

# RESULT 3

US-09-115-746-8  
Sequence 8, Application US/09115746  
Patent No. 6228601

## GENERAL INFORMATION:

APPLICANT: Kirchhoff, Louis V.  
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/115,746  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/216,894  
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399



TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 643 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-115-746-8

Query Match 6.7%; Score 150.5; DB 4; Length 543;  
 Best Local Similarity 21.5%; Pred. No. 0.0015;  
 Matches 99; Conservative 56; Mismatches 169; Indels 135; Gaps 17;

QY 7 PGPIDEFERPPADLSAGLEASANKSAEAOIAGAEAKPKESKTDSEKMSILRSAN 66  
 DB 216 PPKSDLVPRGSPOL--QQAENNTINSKEMTKLEKVKKAKEKLDAINRATKLEEBRN 273  
 QY 67 ALMSLADKLGIASSNSSSTSRADVDSTTATPTPPPTSDDYKTOAYDTFTST 126  
 DB 274 QAYAAHK--AEERKAKTFORLTFESENINLKR-- 307  
 QY 127 LADIAALVSLQDAVTN-----IKDTATDE-----ETAIAAEWETKNADAIKVGAI 174  
 DB 308 -----NDAVSNRDKKKNSETAKTDEVEKQRAAEAKAVETEKQRA-----AFA 350  
 QY 175 TELKAYASDNOAIIDSLGKLTSPDLLOTTALLQSVANNKKAELKEMODNPVPGKTPAI 234  
 DB 351 TKVAE-----AEKRAAEAKAVETEKQRAAEATKV 381  
 QY 235 A-----OSLVDTDTATPIEKDGNAGIDAYFAGONASGAVENAKSN-----NSISMD 283  
 DB 382 AEAEKQRAAEAKAVETEKQRAAEATKVAEKQRAAEAKAVETEKQRAAEATKVAAE 441  
 QY 284 SAKAIAITAKTQIAEAK--KEPDSPIQAEOM-----VIOEKDKNIKPADGSDVP 335  
 DB 442 KQKAAEA---TKVAEAEKQRAAEATKVAEKQRAAEATKVAAEKO---KAAEATVVA 494  
 QY 336 NPGTVGSKQSSIGSIRVSMILDDAENETASTILSGFQMIHMTENPDSDQAQE 395  
 DB 495 E-----AEKQKAAEATKV-----EAEKQKAAEATK-----VAAEAEKQKAAEAT 533  
 QY 396 LAAQARAFAKAGDDSAALADAKALAEALGAKGQOOGIL 436  
 DB 534 KVAEAEKQKAA--EATKVAAEAEKQRAAEAKAVESQKORFL 572

RESULT 4  
 5352450-2  
 Patent No. 5352450  
 APPLICANT: KOGA, TOSHIIKO; OKAHASHI, NOBUO; TAKAHASHI, ICHIRO;  
 SHIBUYA, KOJI; OHYA, HIROTAKA  
 TITLE OF INVENTION: METHOD FOR PREPARING VACCINE FOR DENTAL  
 CAVITIES AND VACCINAL COMPOSITIONS FOR DENTAL CAVITIES USED AS NASAL  
 DROP  
 NUMBER OF SEQUENCES: 2  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/529,602  
 FILING DATE: 29-MAY-1990  
 SEQ ID NO: 2  
 LENGTH: 1565  
 5352450-2

Query Match 6.5%; Score 146.5; DB 6; Length 1565;  
 Best Local Similarity 21.8%; Pred. No. 0.011;  
 Matches 101; Conservative 81; Mismatches 190; Indels 91; Gaps 21;  
 QY 48 KESTDSVERMSILRSANALMSLADKLGIASSN--SSSSSRADVDSTTATPTPP 105  
 DB 2 KVKTTYGFRKSKISKILGAVLGIVAAVSAVGQVFADETTTSDVDTKVVGTGPNFA 61  
 QY 106 T-----SDKYTOAQATAYDTFTST-----SLADIAALVS-LQDAVTNIKD 146

DB 62 TQLPEAGSASKEAEQSOTKLERQVHTIEPKTLDQAAKAKSGVNVQDADVN--KG 120  
 QY 147 TAATDEETAIAAEWE-----TKNADAIK-----VGAQITELAKYASDQALDLSG 192  
 DB 121 TKTTEE-AVOKETIEKEDYTKQADIKTTDQTSVDAHAHEAEVAKKAKKQATKEQYE 179  
 QY 193 -KTSFDLLQTTALLQSVANNKKA--ELKEMODNPVPGKTPAIASQSLVDQTDATQIE 250  
 DB 180 KDMAHRAEVEERINANNAASKTAYEAKLAQYQADLAAYQKTAANQAAV--QALAAAYQAE 238  
 QY 251 ----KDGNAGIDAYFAGONASGAVENAKSNNSISIDSAKAIATAK-----TQIEAQ-- 300  
 DB 239 LKRVQEAANNAKAAV--DTAVANNAKNTETIAANEEIRKNATATKAETETKLAQYQAE 295  
 QY 301 -KEPDSPIQAE--QVIOA--EKDLKNIKPAD-----GSDVPNPGTVGGS- 344  
 DB 296 LKRVQEAANNAEADQALUTAYQTELARQKANAADAKATYEAAVANAANKNAALTAEVTA 355  
 QY 345 -----KQSSIGSIRVSMILDDAENETASTILSGFQMIHMTENPD 388  
 DB 356 IKERNENAKAYEALQYEAALQYEAALQYEAALQYEAALQYEAALQYEAALQYEAALQY 412  
 QY 389 SOAAQOELAAQARAFAK--GDDSAAL--ADAKALAEALGK 428  
 DB 413 KAAATEAAVANAANAALTAEVTAIKRNADAKADYEKILAK 455

RESULT 5  
 US-08-687-956A-23  
 Sequence 23 Application US/0867956A  
 Patent No. 5861157  
 GENERAL INFORMATION:  
 APPLICANT: BURNIE, JAMES P  
 APPLICANT: MATTHEWS, RUTH C  
 TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF  
 NUMBER OF SEQUENCES: 23  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: PILLSBURY, MADISON & SUTRO, LLP  
 STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH  
 CITY: WASHINGTON  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20005  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/687,956A  
 FILING DATE: 29-JUL-1996  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9401689.6  
 FILING DATE: 28-JAN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KOKULIS, PAUL N  
 REGISTRATION NUMBER: 16,773  
 REFERENCE/DOCKET NUMBER: 50885/222892  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202/861-3000  
 TELEFAX: 202/822-0944  
 INFORMATION FOR SEQ ID NO: 23:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1566 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO



ANTI-SENSE: NO  
 FRAGMENT TYPE: Internal  
 ORIGINAL SOURCE:  
 ORGANISM: Streptococcus sobrinus  
 US-08-687-956A-23

Query Match 6.38; Score 143; DB 2; Length 1566;  
 Best Local Similarity 22.8%; Pred. No. 0.019;  
 Matches 94; Conservative 67; Mismatches 179; Indels 72; Gaps 17;

QY 48 KESTDVSERMSILRSANALMSLADKIGIASNSSSSSTS--RSADVSTATATATPPPP 105  
 DB 10 KRSKISRLGALLGTAL--LASVYGKALAEETSTSTGCVNVAVTETGNPATNLPD 67  
 QY 106 TSDYKTOAQATAY-----DITFTSLADIQALVALSLQDAVNTIKDTAETATAIA 158  
 DB 68 KQDNPSSQAEFTSOAQAGKGTGMSVSTSELDEAKSAQEGVTVSQDAFVDKCTVETS 127  
 QY 159 EWEKKNADAK--VGAQTTELAKYASD-----NQALIDSLGKITSFDLQTLALQSYA 209  
 DB 128 DEANQKETEIKDDYSKQADIQKTEDEYKAAVARNOAETD---RITQENNAKKAYEODL 184  
 QY 210 NNNKAELLKEMODNPVVPKTPAISOVLQDTATQIEKQNAIGDAYFAGNASGA 269  
 DB 185 AANK-AEVEKITNEN-----AQKADY-EAKLAQYQKDLAAVOQANNSQAAYAA 232  
 QY 270 VENAKSNSISNTDSAKAIAITAKTQIAE-----QKFPDSPILQ-----EAE 313  
 DB 233 AKEA-YDEKLARVOAANA--AKKEYEALAAVTTKNEQJKAENAAIQORNAQAKADYE 288  
 QY 314 QMVIAEKDLNIRKADSDVNPCTYVGGSGQSSSTGSIKRVSLDDAENETASILMS 373  
 DB 289 AKLQYEDKDLAAOSGNATNEADYQAKRAVEQELARV-----QAAAKA---Q 335  
 QY 374 GFRQIHMFENTPDSQAQOELAAQARAKAGDSSAAAALADAQKALEAA 425  
 DB 336 AVEQALANTAKN--AQITAEENALIQORNAQAKA--NYEAKLAQYQKDLAA 383

RESULT 6  
 US-08-894-2  
 ; Sequence 2, Application US/08216894  
 ; Patent No. 5876734  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kirchhoff, Louis V.  
 ; APPLICANT: Otsu, Keiko  
 ; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
 ; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington, D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/216,894  
 ; FILING DATE: 24-MAR-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)672-5300  
 ; TELEFAX: (202)672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 564 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-216-894-2

Query Match 6.1%; Score 138.5; DB 2; Length 564;  
 Best Local Similarity 21.4%; Pred. No. 0.01;  
 Matches 97; Conservative 54; Mismatches 165; Indels 137; Gaps 17;

QY 7 PGPIDETERTPPADLSAQGLESAANKSAEQRINGAEAKPREKSTDSVERMSILRSANV 66  
 DB 216 PPKSDLVPRGSPSOL--QQAENNTNNSKKEMTKLEKYKKAKEKLDINRATKLEBERN 273  
 QY 67 ALMSLADKLGIASSNSSSTSRADVDSTTATAPPTSDYKTOAQATVITFTSTS 126  
 DB 274 QAYKAHK--AEKAKTFORLITFESEINILKKRP----- 307  
 QY 127 LADIQALVSLQDAVTN-----IKDTAETD-----ETAIAMETKNADAIYGAOI 174  
 DB 308 -----NDAVSNRDKKKNSETAKTDEYKORAAEAAKAVETEKORA-----AEA 350  
 QY 175 TELAKYASDQALIDSLGKITSFDLQTLALQSVANNKKAELLKEMODNPVVPKTPAI 234  
 DB 351 TKVAE-----AEKRAAABAARAVETEKORAAEATKV 381  
 QY 235 A-----QSLVDQDATATQIEKQNAIGDAYFAGNASGAVENAKSNSISNTDSAKAIA 290  
 DB 382 AAEKQKAEEAKKAVETEKORAAEATKVAAEAKQAAEAMK-----VAAEKQKAABA 434  
 QY 291 TAKTQIAEAK-KFPDSPILQEAQOM-----VIOAEKDLNIRKADSDVNPCTYV 342  
 DB 435 ---TVAAEAEKQKAEEATKVAAEAKQAAEATKVAAEAKO---KAAEATKVAE----- 481  
 QY 343 GSKQOSSSTGSIKRVSLDDAENETASILMSGFRQIHMFENTPDSQAQOELAAQARA 402  
 DB 482 AEKQKAEEATKVA-----EAEKQKAEEATKVAEAK 526  
 QY 403 AKAAGDSSAAAA---LADAQALAEALGKAGQ 431  
 DB 527 OKAAEATKVAEAEKQKAEEATKVAAEAKQKAGE 559

RESULT 7  
 US-09-115-746-2  
 ; Sequence 2, Application US/09115746  
 ; Patent No. 6228601  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kirchhoff, Louis V.  
 ; APPLICANT: Otsu, Keiko  
 ; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
 ; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington, D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/115,746  
 ; FILING DATE:  
 ; PRIOR APPLICATION NUMBER:  
 ; APPLICATION NUMBER: 08/216,894  
 ; ATTORNEY/AGENT INFORMATION:



NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 564 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-115-746-2

Query Match 6.1%; Score 138.5; DB 4; Length 564;  
 Best Local Similarity 21.4%; Pred. No. 0.01;  
 Matches 97; Conservative 54; Mismatches 165; Indels 137; Gaps 17;  
 QY 7 PGIDERTPPADLSAOGLEASAAKSAEAQRIAGAEAKPKESKTDVSVERWSILRSVAVN 66  
 DB 216 PPKSGLVPRGSPQL--QOENNITNSKKEMTKLREKVKAEKELDRINRATKLEERN 273  
 QY 67 ALMSLADKLGIASSNSSSSRSADVDSTTATPTPPPTSDDYKTKQAQAYDTFTSTS 126  
 DB 274 QAYKAHK---ABEEKAKTFORLITPESINILKKRP----- 307  
 QY 127 LADIAALVSLQDAVN-----IKDTAATDE-----ETAIAAEWETKNADAIVGCAQI 174  
 DB 308 -----NDVSNRDKKKNSSETAKTDEVEKQRAAEAAKAVETEKORA-----AEA 350  
 QY 175 TELAKYASDNOAILDLSGLKLTSPDLLQTLQSVANNKAAELLLKEMODNPVVPKTPAI 234  
 DB 351 TKVAE-----AEKKAEEAAKAVETEKQRAAEATKV 381  
 QY 235 A-----QSLVQDTATQIEKDGNAIGDAYFAGONASGAVENAKNSNISIDSAKAAIA 290  
 DB 382 AEAKQKAAEAARAEKTEKQRAAEATKVVAEAKQRAAEAMK-----VAEAEKQKAAEA 434  
 QY 291 TAKTOIAEAK-KFPDPIQAEOM-----VIOAEKDLKNIKPADGSDVNPNGTVG 342  
 DB 435 ---TKVAEAEKQRAAEATKVVAEAKQKAAEAATKVVAEAKQ-----KAAEATKVAE- 481  
 QY 343 GSKQGGSIGSIRVMSLLDADENETASILMSGFRQIMHFNENTNPDSQAAQQLAAQARA 402  
 DB 482 AEKQKAAEAATKVA-----EAERKAAEAATK-----VAEAEKQKAAEAATKVVAEAK 526  
 QY 403 AKAGDDSAANA-----LADAKALEAALGRAGQ 431  
 DB 527 QKAAEATKVVAEAKQKAAEAATKVVAEAKQKAGE 559

RESULT 8  
 US-07-741-940-7  
 Sequence 7, Application US/07741940  
 Patent No. 5352775  
 GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS  
 APPLICANT: ANAND, RAKESH  
 APPLICANT: CARLSON, MARY  
 APPLICANT: GRODEN, JOANNA  
 APPLICANT: HEDGE, PHILIP J.  
 APPLICANT: JOSLYN, GEOFF  
 APPLICANT: KINZLER, KENNETH  
 APPLICANT: MARKHAM, ALEXANDER F.  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: THLIVERIS, ANDREW  
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
 NUMBER OF SEQUENCES: 94  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Banner, Birch, McKie & Beckett

STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20001-4598  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA: US/07/741,940  
 APPLICATION NUMBER: US/07/741,940  
 FILING DATE: 19920109  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kagan, Sarah A.  
 REGISTRATION NUMBER: 32,141  
 REFERENCE/DOCKET NUMBER: 1107.035574  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-508-9100  
 TELEFAX: 202-508-9299  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2842 amino acids  
 TYPE: AMINO ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: APC  
 US-07-741-940-7

Query Match 6.1%; Score 138; DB 1; Length 2842;  
 Best Local Similarity 20.0%; Pred. No. 0.11;  
 Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERTPPADLSA--QGLEASAAKSAEAQRIAGAEAKPK-----SKTDSV----- 55  
 DB 1658 DLTIESPNEULAGEGVGGAGQSGEFKEKRTIPIEGRSTDEAOGGKTSSVTIPELDNKA 1717  
 QY 56 ERWSILRSVAVNLM-----SLADKLGIASSNSSSSRSADVDSTTATATPTTP 103  
 DB 1718 EEGDILAEICINSAPKPKSHKPRVKIMDQVQOASASSAPNKNQLDGKKKPTSPVKP 1777  
 QY 104 PPTSDDYKTKQAQAYDT-----IFT-----STSLADIQAALVSLQDAV----- 141  
 DB 1778 IPQNTYRTRVYRNKADSKNNLNAERVFSDNKRDKQNLKNNKDFNDKLPNEDRVGRSF 1837  
 QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVGCAQITELAKYASDNOAILDLSL 191  
 DB 1838 AFDSPHHYTPIEGPPYCFSRNDSLSLDDDDDDVLSREKAE-----RKAKENK-----ESE 1890  
 QY 192 GKLTFSOLLQTLQSVANNKAAELKEM-----QDNVPVPGKTPAIAQSLVDQDTATAT 247  
 DB 1891 AKVTS-----HTELTSNQOSANKTOIAKQPINRGOPKPILO-KOSTFPQSSKDIPIRGAA 1945  
 QY 248 QIEKDGNAIGDAYFAGONASGAVENAKNSNISIDSKAAIAATAKTOIAEAKKFPDPS 307  
 DB 1946 TDEKLQ-----FAIENTPVCFSHNSLSLSLSDID----- 1975  
 QY 308 ILOEAQEMVIOAEKDLKNIKPADGSDVNPCTTVGGSKQ-----QGSSIGSIR 355  
 DB 1976 --QENNN--KENEPKETETEPDQSGEPKQASGYAPKSFHVEDTPVCFERNSSLSLS 2030  
 QY 356 VSMLLDDAENETASILMSGFRQIMHFNENTNPDSQAAQQLAAQARAAGDSDSAAL 415  
 DB 2031 I-----DSEDDLQECISS-----AMPKKKKPSRLKGNKHSPPNMGILGED-LTLDL 2079  
 QY 416 ADAQKA-LEAALG-----KAGQQ--QGLNALGQIASAAVVS 449



Db 2080 KDIQPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAAAAACLS 2124

RESULT 9

US-08-289-548A-7

Sequence 7, Application US/08289548A

Patent No. 5648212

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Allegretti, LTD

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/289,548A

FILING DATE: 12-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.46943

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2842 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: APC

US-08-289-548A-7

Query Match 6.18; Score 138; DB 1; Length 2842;

Best Local Similarity 20.08; Pred. No. 0.11;

Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

Qy 11 DETERTPADLSA-OGLEASAAKSAEAOIAGAPKPE-----SKTDSV----- 55

Db 1658 DLTIESPPELAAEGVGGAGQSGEFERDPTPEGRSTDQAQGGKTSVTPPELDONKA 1717

Qy 56 ERWSTLRSAVNAIM-----SLADKLGIASSSSSTSRSDVDSTTATPTPP 103

Db 1718 EEGDILAEICINSAMPKGSKHPKRVKIMDQVQOASASSAPNQLDGGKKKPTSPVKP 1777

Qy 104 PPTSDDYKTAQATYDT-----IFT-----STSLADLQAALVSLQDAV----- 141

Db 1778 IPONTEYTRVRKNADSKNNLNAERVFSNDKSKQNLKNNKSKDFNDKLPNNEORVRGSG 1837

Qy 142 -----TNIKDT--AATDEETAAAEWETKNADAIVKGAQITELAKYASDNQALDLSL 191

Db 1838 AFDSPHHYPIEGTPYCFRNDLSLDDDDVDLSRKAEL-----RKAKNK-----ESE 1890

Qy 192 GKLTSPDLLQTTALLOSANNKAAELLKEM---QDNVVPVPGKTPAIAQSLVDOTDATAT 247

Db 1891 AKVTS---HTELTSNQOSANKTQATAKOPINRGQPKILO-KOSTFPOSSKDIPDRGAA 1945

Qy 248 QIEKQGNATGDAYFACQNASGAVENAKSNNSISNIDSAKAAATATQTAEAKKKFPDSP 307

Db 1946 TDEKLQN-----FAIENTPVCFSHNSLSLSDID----- 1975

Qy 308 ILQEAQOMVIOAEKDLKNIKPADGSDVPNPVGTTVGSGKQ-----QGSSIGSIR 355

Db 1976 --QENNN---KENEPKETEPPDSQGEPSKQASGVAPKSFHVEDTPVCFSRNSLSLSL 2030

Qy 356 VSMLLDDAENETASILMSGFRQIMHMTENTENPDQSOAAQOELAAQAAKAAAGDSDSAAAL 415

Db 2031 I-----DSEDDLQECISS-----AMPKKKKPSRLKGDNEKHSRNNMGILGED-LTLDL 2079

Qy 416 ADAQKA-LEAALG-----KAGQQ--OGILNALGOIASAAVVS 449

Db 2080 KDIQPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAAAAACLS 2124

RESULT 10

US-08-452-654-7

Sequence 7, Application US/08452654

Patent No. 5691454

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner, Birch, McKie & Beckett

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/452,654

FILING DATE: 25-MAY-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/741,940

FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.035574

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2842 amino acids